Three-Dimensional Reconstruction of *Heterocapsa circularisquama* RNA Virus by Cryo-Electron Microscopy

By Jennifer Louise Miller

A minithesis submitted in partial fulfilment of the requirements of the degree of Master of Science (Structural Biology) at the University of the Cape Town.

Supervisors:
Associate Professor Trevor Sewell
Dr Arvind Varsani

November 2007

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Keywords

Heterocapsa circularisquama RNA virus

Icosahedral virus

ssRNA virus

Electron microscopy

Structure

3D reconstruction

Common lines

CTF

Abstract

Three-Dimensional Reconstruction of *Heterocapsa circularisquama* RNA Virus by Cryo-Electron Microscopy

By J. L. Miller

MSc minithesis, Faculty of Science, University of Cape Town and Faculty of Natural Sciences, University of the Western Cape.

Heterocapsa circularisquama is a bloom-forming, bivalve-killing dinoflagellate that is responsible for causing red tide around Japan and Hong Kong. It is infected by two viruses: *H. circularisquama* virus (HcV) and *H. circularisquama* RNA virus (HcRNAV). HcRNAV, a non-enveloped icosahedral ssRNA virus, is the major natural agent controlling the population dynamics. While there are many strains of the virus, they can be classified into two types, UA and CY, based on their intra-species strain-specific infectivity. The difference in specificity has been shown to be limited to the entry process of the virus into the host cell, presumably due to variations in the sequence of the capsid proteins.

This study commences the exploration for the structural explanation of the strain-specificity by determining the three dimensional structure of HcRNAV109, a CY-type HcRNAV strain, by both cryo-electron microscopy and negative stain electron microscopy.

The virus was reconstructed by single particle reconstruction methods from two independent starting models. The first model was produced by using common lines methods to determine the orientations of rotationally invariant *K*-means clustered class averages while the second starting model was a featureless model with icosahedral symmetry. Refinement of the models was carried out in separate defocus groups for the cryo-EM data set and in one group with no CTF correction for the negative stain data set. Projection matching was iteratively used to refine the orientations of the images and produce improved reconstructions. For the cryo reconstruction the calculated CTF was applied to the initial model for each defocus group so that the images and the references to which they were aligned have the same CTF. The projections of this model were then used as references for multi-reference alignment of the images to determine the orientations of the particles which were back-projected to form an improved model. All the models from the different defocus groups were then combined with CTF correction to form one CTF-corrected model that was used for the next iteration.

A final resolution of 18.1 Å, computed by comparing the FSC of the two independent reconstructions, reveals that the virus is a T=3 icosahedral virus consisting of 180 protein subunits, contrary to what was predicted by previous homology modelling to black beetle virus. While the negative stain model only reveals the surface features, the model produced from the cryo-EM data reveals some of the internal RNA structure which is generally disordered, except for around the three-fold axes where the RNA binds to the capsid protein. Further studies of the UA-type virus and higher resolution studies are necessary to determine the effect of the sequence differences on the structure of the capsid protein.

November 2007

Declaration

I declare that ***Three-Dimensional Reconstruction of Heterocapsa circularisquama RNA Virus by Cryo-Electron Microscopy*** is my own work, that it has not been submitted for any degree or examination in any other university, and that all sources I have used or quoted have been indicated and acknowledged by complete references.

Jennifer Louise Miller 21 November 2007

Acknowledgements

Firstly, I would like to thank my supervisors. Trevor, thank you for your wisdom and advice during this project and all you have taught me during the course work.. Arvind, for your tireless proofreading, writing advice and virus knowledge, I am very grateful.

Thank you to Keizo Nagasaki for providing the virus sample. Many thanks to Alan Roseman, Shaoxia Chen and Tony Crowther from the MRC Laboratory of Molecular Biology for all of their help and advice.

I am grateful to Brandon and Mohammed for their teaching of laboratory techniques and electron microscopy and all their patience. Many thanks to the rest of the staff of the EMU for all of their help too.

Thank you to Leandri for the sanity-maintaining coffee breaks and amazing friendship, to Johann for the support and friendship over these two years, to Jeremy, Kyle, Jason and Ndoriah for the structural biology conversations and other distractions.

I would like to thank my family and friends for their support and encouragement over the years.

I gratefully acknowledge the financial support of the National Research Foundation and the Carnegie Corporation of New York.

List of Abbreviations

2D	Two-dimensional
3D	Three-dimensional
BBV	Black beetle virus
CTF	Contrast transfer function
DNA	Dioxyribonucleic acid
DPR	Differential phase residual
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA
EM	Electron microscopy
EMBL	European Molecular Biology Laboratory
FEG	Field emission gun
FSC	Fourier shell correlation
HAC	Hierarchical ascendant classification
HcRNAV	*Heterocapsa circularisquama* RNA virus
HcV	*Heterocapsa circularisquama* virus
LMB	Laboratory of Molecular Biology
MDA	Multivariate data analysis
MRC	Medical Research Council
MSA	Multivariate statistical analysis
ORF	Open reading frame
RdRp	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
SNR	Signal-to-noise ratio
SPIDER	System for Processing Image Data from Electron microscopy and Related fields
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
T-number	Triangulation number
TEM	Transmission electron microscopy

Table of Contents

Chapter 1. Introduction

1.1. Heterocapsa circularisquama RNA virus

"Red tide" is the common name for algal blooms, a phenomenon caused by the rapid growth and bloom of marine, estuarine or fresh water algae. The water acquires a red, brown or green colour due to the accumulation of phytoplankton, macroalgae or colourless heterotrophic protists. The algae are dinoflagellates, single-celled, aquatic organisms with two dissimilar flagella, that release toxins into the water killing shellfish and other sea-life (Nagasaki K. *et al.*, 2006; Sellner *et al.*, 2000). Dinoflagellates are either photosensitising autotrophs, endocytotic feeding heterotrophs, or a combination of the two (Nagasaki K. *et al.*, 2006).

Heterocapsa circularisquama, shown in Figure 1.1A, is a small dinoflagellate with a case covering called theca, that was first observed as the cause of large-scale red tide blooms occurring at Uranouchi Bay, Kochi Prefecture, Japan in August 1998. It was subsequently observed in other areas of western Japan (Matsuyama, 1999) and off the coast of Hong Kong (Iwataki *et al.*, 2002). This bivalve-killing, bloom-forming dinoflagellate was first described by Horiguchi (1995) using samples collected at Ago Bay, Mie Prefecture, central Japan in 1992. While exhibiting no harmful effects on fish, it contains a photo sensitising hemolytic toxin which is lethal to shellfish and some microzooplankton (Miyazaki *et al.*, 2005).

H. circularisquama virus (HcV; ~200 nm; icosahedral DNA virus; Tarutani *et al.*, 2001) and *H. circularisquama* RNA virus (HcRNAV; 30 nm; icosahedral RNA virus; Tomaru *et al.*, 2004) were the first two dinoflagellate-infecting viruses that were characterised. HcRNAV is the major natural agent controlling the *H. circularisquama* population dynamics, infecting 88% of the dinoflagellates at the peak of the bloom in Ago Bay, Japan in 2001. As HcRNAV is found in sediments, with the abundance decreasing more gradually than in the water, the sediments are thought to possibly be a reservoir for the virus, which is an important factor affecting the population dynamics of *H. circularisquama* (Nagasaki *et al.*, 2004).

H. circularisquama cells when grown in culture and infected with HcRNAV lose motility and settle. During early stages of infection the intracellular structures are mostly intact, with cytoplasm structure and organelle degradation only clearly visible at 33 to 48 hours post-infection, during which time newly formed virus particles appear in the cytoplasm (Figure 1.1B). The viruses form either crystalline arrays (Figure 1.1C) or unordered aggregations in the cytoplasm. Cell numbers

decrease 24 to 48 hours after infection, due to lysis, however this is not complete as 10 to 20% survive, although they are immotile and lose their theca (Tomaru *et al.*, 2004).

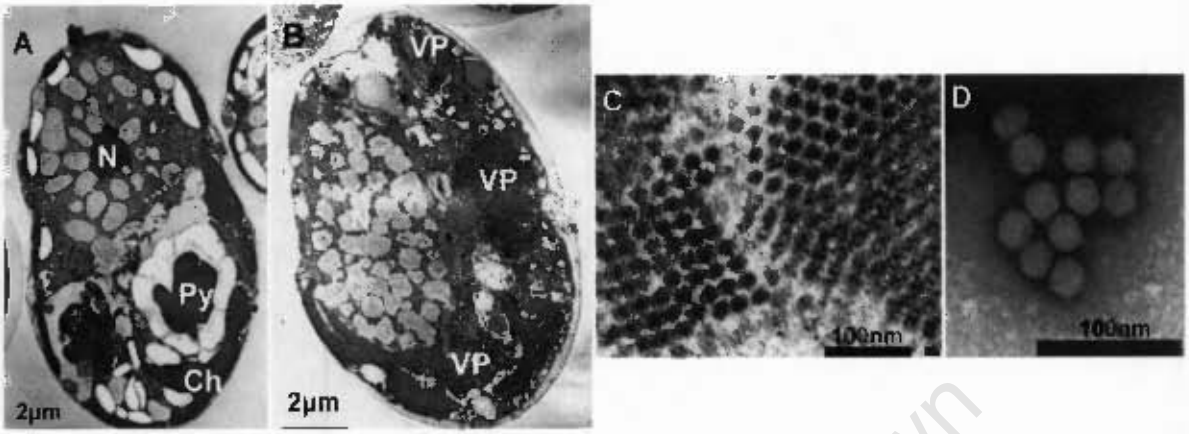

Figure 1.1: *Transmission electron micrographs of HcRNAV109 infected H. circularisquama (HCLG-1 strain). **A)** Healthy HCLG-1 H. circularisquama cell showing the nucleus (N), chloroplasts (Ch) and pyrenoids (Py). **B)** 48 h after infection, showing viroplasms (VP) and degradation of organelles in the cytoplasm. **C)** Crystalline array formation of HcRNAV109 in the cell. **D)** Negatively stained HcRNAV109 isolated from the lysate (Tomaru et al., 2004).*

HcRNAV is strain-specific as to which marine plankton it infects. It was found not to be lytic to the 32 marine phytoplankton strains, including 11 strains of the genus *Heterocapsa*, that were tested by Tomaru *et al.* (2004). Additionally, different strains of HcRNAV will complementarily infect different *H. circularisquama* strains. The virus is thus ecotype-, rather than species-, specific (Tomaru *et al.*, 2004). HcRNAV can be classified as UA-type or CY-type based on their intra-species strain-specific infectivity. *H. circularisquama* strains can consequently be classified into three types: those that are lysed by HcRNAV UA-type, those lysed by HcRNAV CY-type, and those that exhibit no lysis for either type of HcRNAV (only two of the 56 strains tested; Tomaru *et al.*, 2004). The two host/virus systems are independent and may coexist in a bloom, thereby affecting both the quality (*i.e.*, the clonal composition) and the quantity (biomass) of the *H. circularisquama* bloom (Tomaru *et al.*, 2007).

Two strains, HcRNAV34 and HcRNAV109, have been characterised as typical type UA and CY virus strains. These nonenveloped, spherical viruses with no tail structure, contain a single-stranded RNA genome approximately 4.4 kb long (4.375 kb for HcRNAV34 and 4.391 kb for HcRNAV109). The genomes of HcRNAV109 and HcRNAV34 are 97% homologous and both have

two major open reading frames (ORFs). In addition there are four and six minor ORFs that could code for proteins of approximately 100 residues for HcRNAV34 and HcRNAV109 respectively. The function of these small ORFs is yet to be determined as they showed no significant BLAST similarity to any known proteins (Figure 1.2; Nagasaki *et al.*, 2005).

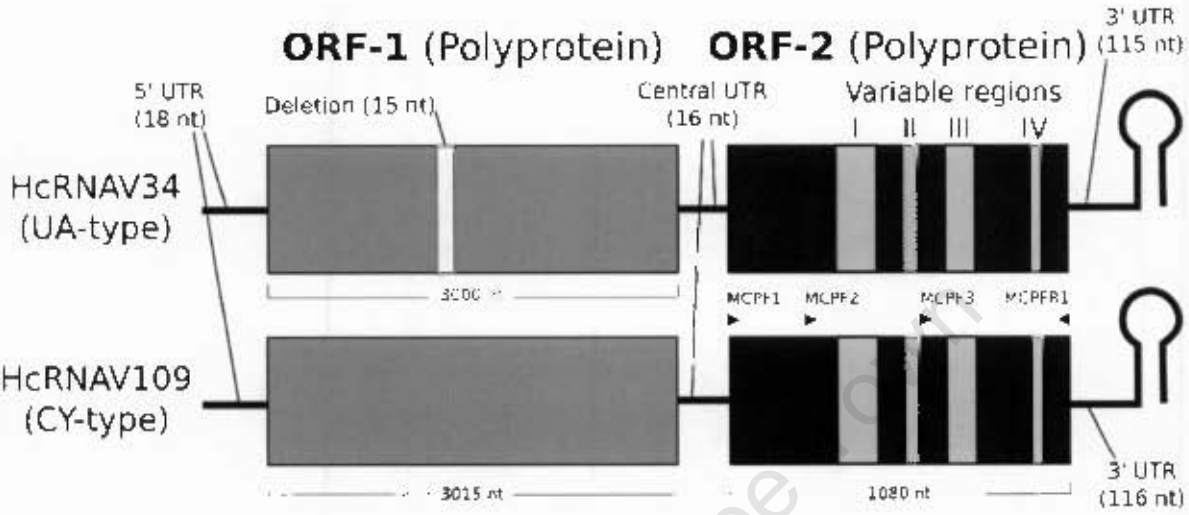

Figure 1.2: *Schematic genome structure of HcRNAV34 and HcRNAV109 and the positions of the variable regions. Based on Nagasaki et al. (2005).*

ORF-1 (15 nucleotide size difference between the two strains) has two conserved domains: a serine protease domain and an RNA-dependent RNA polymerase (RdRp) domain. This is most likely a polyprotein that is cleaved into the two functional proteins. There is only low homology with the serine protease and RdRp domains of other viruses, showing that HcRNAV is quite evolutionarily distant from any of the viruses that have been previously characterised (Nagasaki *et al.*, 2005). This is probably because the database of aquatic viruses is small with HcRNAV being the only characterised virus infecting Dinozoa, a classification that includes dinoflagellates (Nagasaki K. *et al.*, 2006).

The 1080 nucleotides of ORF-2 code for the 38.2 and 38.3 kDa single major structural protein of HcRNAV34 and HcRNAV109 respectively (Nagasaki *et al.*, 2005). The most striking differences between the genomes of the two types are four variable regions in ORF-2 (Figure 1.2). Phylogenetic analysis of the variable regions from a number of strains coincides with the host-specificity of the viruses, with type UA and CY virus strains divided into distinct groups. This suggests that the

specificity is due to the sequence variation in the capsid protein (Nagasaki *et al.*, 2005). This was confirmed by the same group using particle bombardment (Mizumoto *et al.*, 2007).

Mizumoto *et al.* (2007) reported that when virus suspension was added directly to a host culture, the virus infected the cells and was replicated only when the strains were compatible. However, when the cells were infected by transfecting viral RNA directly into host cells by particle bombardment, infective virus particles of the type encoded by the viral RNA were produced irrespective of the host strain, although the RNA accumulation levels did differ across different host-virus combinations. This shows that the host specificity is determined at the entry process, rather than during the intracellular replication processes (Mizumoto *et al.*, 2007).

Nagasaki *et al.* (2005) have modelled the HcRNAV capsid protein based on ~10% sequence homology to the black beetle virus (BBV) capsid protein (Wery *et al.*, 1994) to determine where the amino-acid substitutions occur. The capsid plate is modelled as a trimer of ORF-2 products forming a T=1 icosahedral virus (see Chapter 1, section 1.2.1 Triangulation numbers). Seven and eight out of the 14 significantly different residues were predicted to be on the outside of the UA and CY capsid proteins respectively with no significant amino acid substitutions on the interior of the capsid protein (Nagasaki *et al.*, 2005). The secondary structure prediction of HcRNAV109 capsid using PSIPRED (Jones, 1999), and alignment of the sequences using CLUSTAL X (Thompson *et al.*, 1997) displayed in Figure 1.3 reveals very limited predicted secondary structure homology. Because there are no homologues to the HcRNAV capsid protein, the tertiary structure prediction may not be accurate and therefore structural studies of HcRNAV are necessary in order to assess the validity of these claims.

Figure 1.3: *Predicted secondary structure (using PSIPRED; Jones, 1999) and sequence of the HcRNAV109 capsid protein aligned to the BBV capsid protein sequence and secondary structure.*

1.2. Icosahedral viruses

The structures available for a virus are limited by their need for efficiency. Viruses consist of a high molecular weight nucleic acid component, surrounded by a protective protein packaging, the viral capsid. The nucleic acid, RNA or DNA, encoding for the structural proteins of the virus and sometimes other enzymes necessary for viral replication, must be accommodated inside the capsid. It is therefore advantageous for the capsid to be composed of a large number of identical smaller units as this can be encoded in the nucleic acid more concisely. There are thus only a limited

number of genetically economical basic shapes for a virus: helical tubes and icosahedrons (Caspar & Klug, 1962). HcRNAV falls into the latter group.

An icosahedron is an economical shape for a virus as it consists of 60 asymmetric subunits which allow the virus to form a large shell using minimal amounts of information (Caspar & Klug, 1962). Klug and Caspar (1960) differentiate between three types of subunits: the chemical subunit is the individual protein molecule, the crystallographic subunit is the smallest asymmetric unit, and the morphological subunit is one of the "bumps" or "knobs" of the virus. An icosahedron is a platonic solid composed of 20 equilateral triangles. This highly symmetric solid (point group 532) consists of 6 five-fold axes through the 12 vertices, 10 three-fold axes through the 20 triangular faces and 15 two-fold axes halfway through the 30 edges, producing a total of 60 symmetry elements (Figure 1.4; Baker *et al.*, 1999).

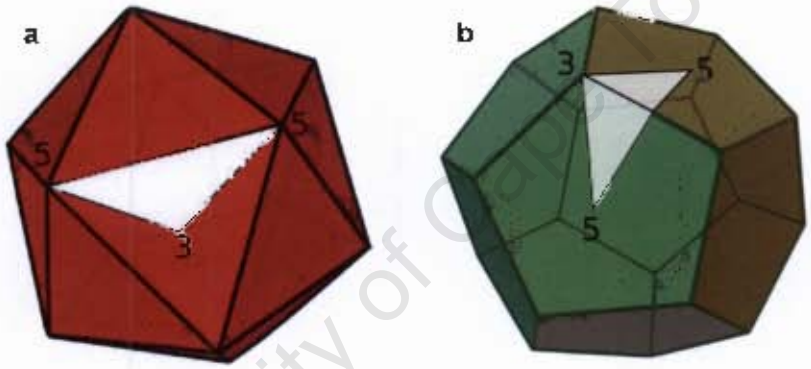

Figure 1.4: ***a)*** *An icosahedron and* ***b)*** *a dodecahedron showing the symmetry axes and the asymmetric unit used in image reconstruction. The numbers (2, 3, and 5) indicate the positions of some of the symmetry axes. The white triangles define the asymmetric unit which is bounded by the lines joining adjacent five-fold and three-fold axes (based on Baker et al., 1999).*

The dodecahedron is related to the icosahedron in that it has the same icosahedral symmetry but it is made up of 12 regular pentagons. The 6 five-fold axes are through the 12 pentagonal faces, the 10 three-fold axes are through the 20 vertices and the 15 two-fold remain through the edges (Figure 1.4). A spherical virus generally has a shape that lies in between the these two icosahedral extremes (Baker *et al.*, 1999).

The asymmetric unit of the icosahedron, which is the $1/60^{th}$ of the total model needed to generate the rest, may be arbitrarily chosen, but is usually the wedge shaped volume enclosed by two adjacent five-fold axes and a three-fold axis as shown in Figure 1.4 (Baker *et al.*, 1999).

1.2.1. Triangulation numbers

While an icosahedral virus always contains 60 asymmetric units, it may be composed of more than 60 protein molecules in order to increase the size of the virus capsid without increasing the size of the genome. The subunits are then arranged in 60 sets of proteins molecules. In this case, although the proteins are chemically identical, their environments and structures differ slightly and thus they are described as *quasi-equivalent* (Caspar & Klug, 1962). The number of quasi-equivalent subunits which make up an asymmetric unit is the triangulation (T) number, which also describes the arrangement of these subunits.

An icosahedron may be constructed from a planar hexagonal net by removing a side from some of the hexagons to form pentagonal vertices in order to generate a closed structure. In the simplest case, adjacent hexagons are bent to form a T=1 icosahedron with all the protein subunits equivalent. A larger icosahedron consists of hexagons between the pentagons and is given a triangulation number $T = h^2+hk+k^2$, where h and k refer to the number of hexamer subunits between pentamers in two directions (Figure 1.5). Therefore a T=3 virus, where $h,k = 1$, consists of a movement up and along one hexamer to move between pentamers.

The number of different environments should equal the T-number with the number of subunits then $60T$, but this is not always strictly maintained (Baker *et al.*, 1999), for example the T=7 papovaviruses have 72 pentamers and six subunit environments, but only 360 subunits (Belnap *et al.*, 1996). This illustrates that viruses do not always strictly adhere to the principle of quasi-equivalence which depends on small conformational and bonding differences.

Some ssRNA viruses that are composed of more than one type of capsid protein, such as comoviruses and picornaviruses, may be classified as P=3 or pseudo-T=3 viruses (Rossmann & Johnson, 1989). Rather than three chemically identical proteins with slightly different environments making up the asymmetric unit, the three proteins differ in their primary structure but have similar tertiary structures (Rossmann & Johnson, 1989). Whereas picornaviruses are composed of three

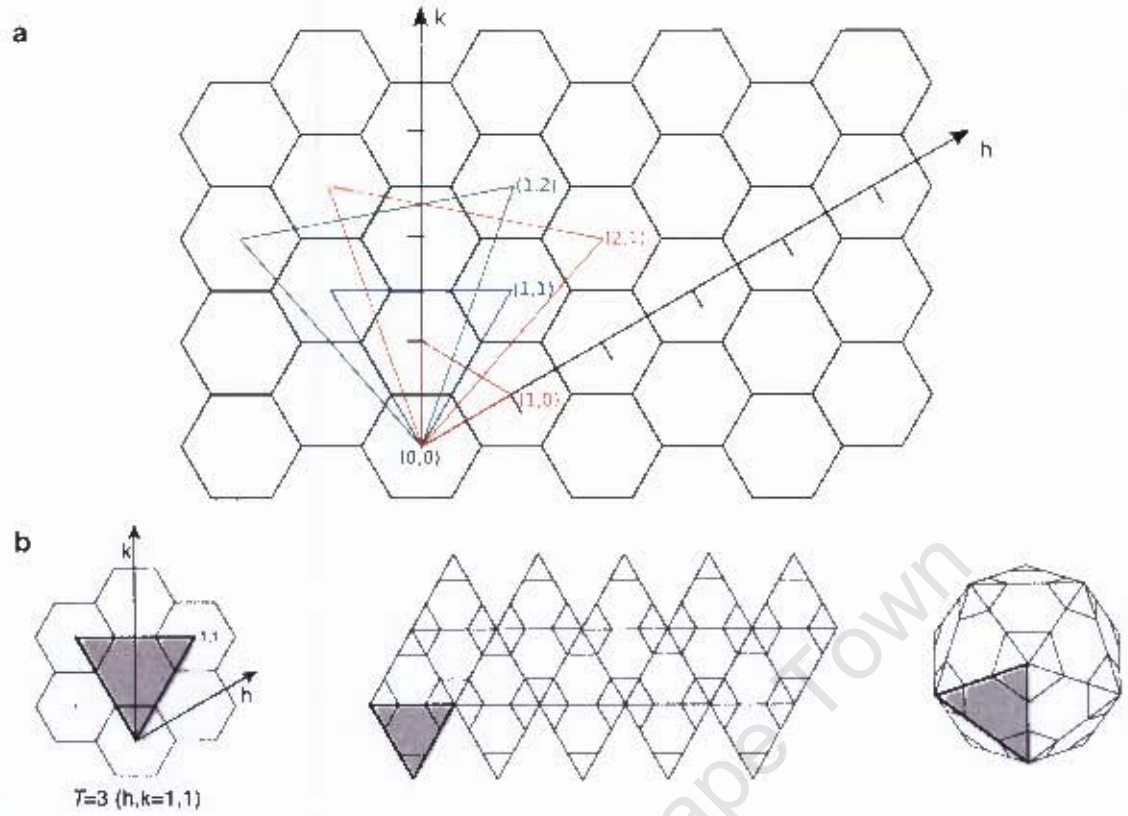

Figure 1.5: *Illustration of the geometrical principle behind T-numbers.* ***a)*** *The hexagonal net with the h- and k-axes marked. To form an icosahedron the certain hexagons are converted into pentagonal vertices. The relative positions of the vertices, identified by their (h, k) coordinates determines the triangular face of the icosahedron and the T-number* ($T=h^2+hk-k^2$). ***b)*** *For example, for T 3 the triangular face of the icosahedron is drawn from (0,0) to (1,1) to (-1,-1). These faces make a net with hexagons converted into pentagons by removing an edge. This can then be folded into a 3D icosahedron (Baker et al., 1999; Johnson & Spier, 1997).*

different protein subunits, comovirus capsids consist of two capsid proteins, L with two domains of similar tertiary structure and S with one domain (Rossmann *et al.*, 1985). This is therefore a T=1 virus, but the structural similarity of subunits and their arrangement leads to a very similar structure to T=3 viruses (Prasad & Prevelige, 2003).

1.2.2. Structure of single-stranded RNA icosahedral viruses

Although the symmetry limits some of the variety in icosahedral viruses, there is still a large diversity in their structures. Spherical viruses consist of an icosahedral protein capsid that may be nonenveloped or enveloped by a lipid membrane. The genetic information contained in the virus

may be in the form of RNA or DNA, which is either single- or double-stranded. Single-stranded RNA (ssRNA) or DNA (ssDNA) viruses may further be classified into positive- and negative-strand depending on whether the nucleic acid has the same (positive) or opposite (negative) sense as the mRNA which can be translated (Dimmock *et al.*, 2001). The positive-strand ssRNA ((+)ssRNA) viruses, which include HcRNAV as well as other viruses from plants, bacteria and animals, are better characterised than the negative-strand ones (Prasad & Prevelige, 2003). In contrast to the negative-strand ssRNA, where the genome is covalently linked to the nucleocapsid proteins and other accessory proteins, there are usually only non-covalent interactions between the naked genome of (+)ssRNA (Prasad & Prevelige, 2003), although the picornaviruses are one of the exceptions as the RNA is bound to a small protein called VPg (Kitamura *et al.*, 1981).

The most common fold for the capsid protein, particularly RNA viruses (Rossmann & Johnson, 1989) is the jelly-roll β-barrel fold (Figure 1.6). It consists of two four-stranded β-sheets that face one another to form a barrel structure that is almost a closed circle. The anti-parallel β-strands tend to lie roughly tangential to the surface of (+)ssRNA viruses and many others, radiating out from the five-fold or quasi-six-fold axes, but in the adenovirus (dsDNA virus) they lie radially (Chapman & Liljas, 2003). Often the jelly-roll proteins have partly disordered N- and C-terminal arms. The disordered segments may interact with the nucleic acid molecule while the ordered parts are used to regulate the subunit packing and stability (Chapman & Liljas, 2003).

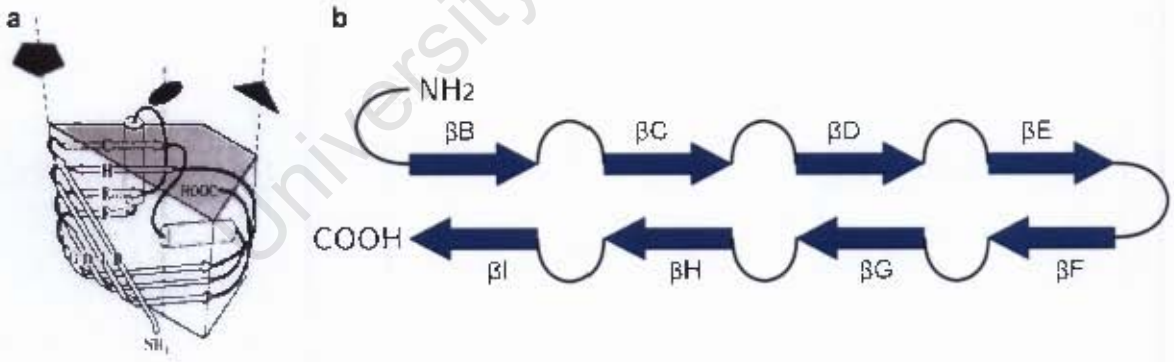

Figure 1.6: *Diagrammatic representation of* ***a)*** *the β-jelly roll subunit and* ***b)*** *topology of the viral subunit. The insertions between the β-strands vary between different viruses. Adapted from Johnson (1996) and Rossmann & Johnson (1989).*

Other folds which are more common in nonviral proteins but are found in some virus structures are the immunoglobulin fold (observed in the flavivirus envelope glycoprotein; Rey *et al.*, 1995), the serine protease fold seen in the Sindbis virus core protein (Choi *et al.*, 1991) and the four-helix bundle that occurs in the helical tobacco mosaic virus (Bloomer *et al.*, 1978). While the jelly-roll, immunoglobulin and serine protease folds all consist of antiparallel β-barrels, the order of the strands and their connections differ (Chapman & Liljas, 2003).

The similarity of structure despite absence of sequence similarity found in many virus structures provides a tool for determining long-range evolutionary relationships between viruses where the sequences have changed beyond recognition but the structure and function is conserved (Bamford *et al.*, 2002). Similarities in capsid protein arrangement as well as the protein folds have led to the recognition of unexpected evolutionary links, such as adenovirus and bacteriophage PRD1 that infect vertebrates and Gram-negative bacteria respectively (Benson *et al.*, 1999).The ubiquity of the β-jelly roll makes it difficult to distinguish between the possibilities that all viruses containing the fold evolved from a common ancestor, or that the fold arose independently in different viral lineages (Bamford *et al.*, 2002).

1.2.3. Single particle reconstruction of icosahedral particles

Whereas X-ray crystallography provides detailed, high resolution structural information about viruses and other molecules, it relies on the availability of large amounts of highly purified sample and the successful growth of diffraction-quality crystals. Electron microscopy (EM) is not limited by these requirements as sample heterogeneity may be overcome during particle picking and may be used to reconstruct virus particles at different stages of maturations (Lee & Johnson, 2003). The structure of the virus produced by EM is also not constrained by crystal contacts. With improvements in reconstruction methods, three-dimensional (3D) reconstruction from EM data can now be used to produce 3D models at subnanometre resolutions at which it is possible to determine secondary structure (Böttcher *et al.*, 1997b; Conway *et al.*, 1997; Zhou *et al.*, 2001). While initially EM was simply used to obtain information about the capsid morphology of the virus (Crowther *et al.*, 1970a), more detailed structural information has been revealed by EM as the methods have improved. Examples of this are the organisation of nucleic acids in viruses (Böttcher & Crowther, 1996), the interaction of antibodies and receptors with the capsid proteins (Prasad *et al.*, 1990), the mechanisms for virus capsid assembly (Bamford *et al.*, 2001; Zlotnick & Stray, 2003) and virus

maturation (Roseman *et al.*, 2005). As previously mentioned structure also highlights interesting evolutionary links between viruses (Bamford *et al.*, 2002).

Transmission electron microscopy (TEM) involves a high-energy (100 keV or higher) electron beam passing through a vacuum to interact with a specimen causing the electrons to be scattered. Widely scattered electrons are screened out by the objective lens aperture to produce amplitude *contrast*, while *phase contrast* (more important in single particle reconstruction) is formed by the interference of the unscattered and elastically scattered electrons. Electromagnetic lenses focus and magnify the electron beam, to form an image on a fluorescent screen, photographic film or a digital camera (Frank, 1996). The image that is recorded is a projection of the sample, with image intensity related to electron density of the specimen. A limiting factor is the sensitivity of the biological material to the radiation, which limits the strength of the electron beam that can be used that would otherwise increase the signal-to-noise ratio (SNR) of the images. While the image is produced by electrons that have interacted elastically with the specimen and subsequently continued to the detector, some electrons interact non-elastically with the specimen, damaging the specimen and thus decreasing the quality of the image as more electrons pass through it (van Heel *et al.*, 2000).

The solution to this problem is to decrease the electron dose using "low-dose" methods for data collection (see Chapter 2). However, this leads to a low SNR, particularly in the high-frequency range (van Heel *et al.*, 2000). Cooling the specimen to liquid nitrogen or liquid helium temperatures protects the specimen from radiation damage allowing for electron doses of about 10 – 15 electrons.$Å^{-2}$ (Thuman-Commike & Chiu, 2000) but this still leads to very noisy images. Therefore it is necessary to average a large number of particle images to increase the SNR and thus improve the reconstruction.

Single particle reconstruction methods are based on the principle that biological macromolecules are identical and thus images of many particles may be combined to produce one 3D model of the particles. Since the particles are randomly orientated, many different views are captured on one micrograph. Determining the orientations of the particles from their two-dimensional (2D) projections and producing a 3D reconstruction from them, is the challenge of single particle reconstruction. Icosahedral particles are well suited to single particle reconstruction methods due to their high symmetry and larger size which produce images containing more information per particle than smaller molecules with less symmetry (Henderson, 2004).

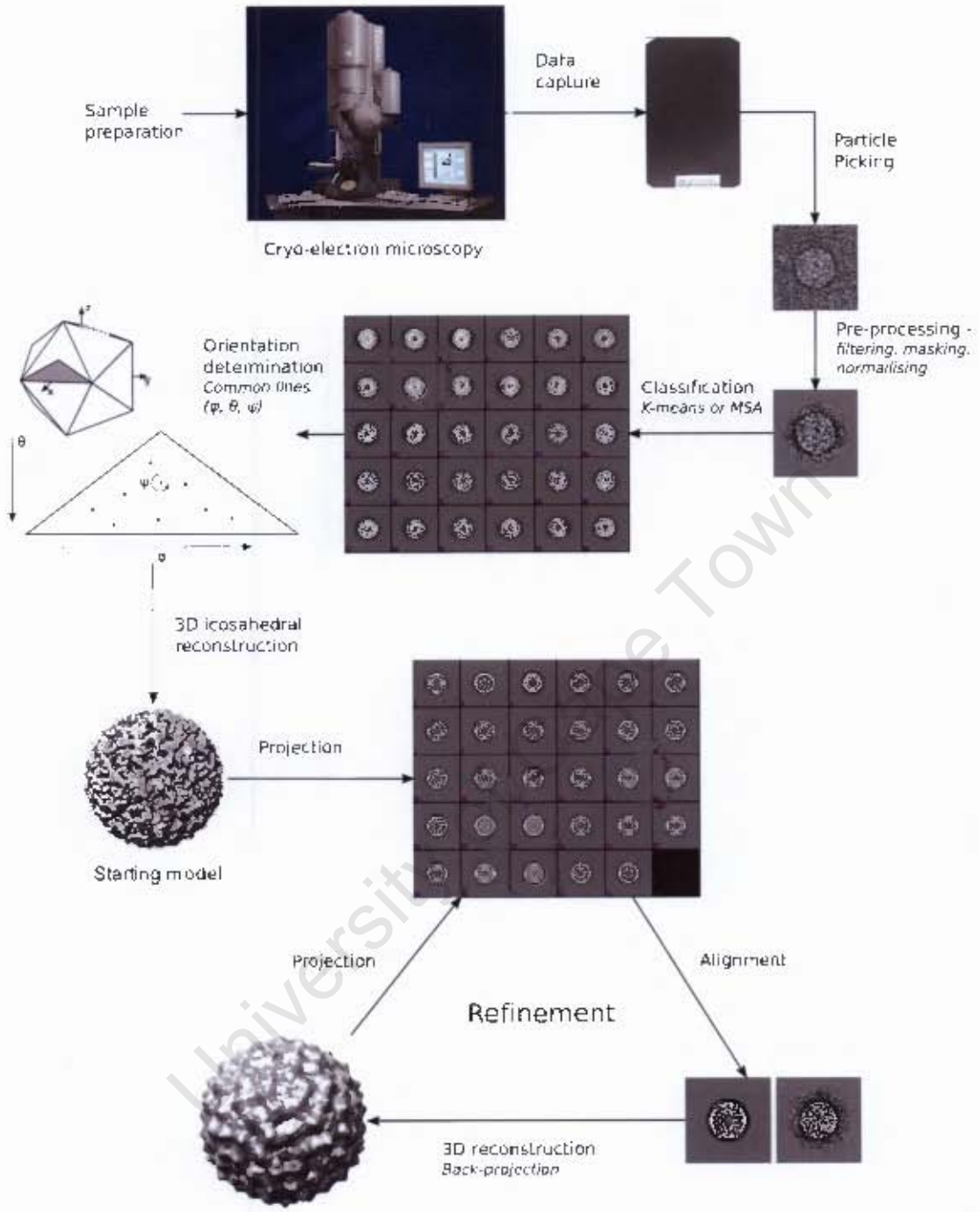

Figure 1.7: *An overview of a method used for single particle reconstruction of icosahedral particles. The isolated and purified sample is inserted into the cryo-electron microscope which is used to capture images of the virus particles at high resolution. These data are digitised and particles are picked from the micrographs in square boxes. Preprocessing of the particles removes information that is unnecessary for further processing. The images are classified in order to determine their orientations so that a starting model may be produced. This is used to iteratively align particles to the projections of the model in order predict an improved model.*

Model-based reconstruction methods consist of iteratively refining the orientations the images of the particles against the projections of model and using the new orientations to produce an improved model. Figure 1.7 outlines one such method, although there are many variations.

The first details about the icosahedral nature of spherical viruses were discovered by Caspar (1956) using X-ray crystallography of tomato bushy stunt virus and turnip yellow mosaic virus. While the first negative stain electron micrographs of icosahedral viruses were taken in 1959, it was not until Crowther's development of the common lines theorem to produce a 3D reconstruction from the 2D EM images of human wart virus (now called human papilloma virus, HPV; Crowther *et al.*, 1970a) and tomato bushy stunt virus (Crowther, 1971), that 3D icosahedral structures could be solved by EM. Some of the viruses that have been solved by cryo-EM are listed in Table 1.1. Many of these are deposited in the Virus Particle Explorer (VIPER) database (Shepherd *et al.*, 2006) and the Electron Microscopy Databank (Tagari *et al.*, 2002).

Table 1.1: *A selection of viruses with structures solved by cryo-EM.*

Virus Name	**Resolution (Å)**	**Year**	**Reference**
Adeno-associated virus 2	10.5	2001	(Kronenberg *et al.*, 2001)
Adenovirus	6.9	2006	(Saban *et al.*, 2006)
Aura Virus (bottom component)	21	2002	(Zhang *et al.*, 2002)
Aura Virus (top component)	17	2002	(Zhang *et al.*, 2002)
Bacteriophage epsilon15	9.5	2006	(Jiang *et al.*, 2006)
Bacteriophage K1-5 (complete)	17	2007	(Leiman *et al.*, 2007)
Bacteriophage K1E (complete)	17	2007	(Leiman *et al.*, 2007)
Bacteriophage P22	20	2006	(Lander *et al.*, 2006)
Bacteriophage phi29 fibered (isometric particle)	8.7	2005	(Morais *et al.*, 2005)
Bacteriophage phi29 fiberless (isometric particle)	7.9	2005	(Morais *et al.*, 2005)
Bacteriophage phi29 fiberless prohead particle	12.7	2005	(Morais *et al.*, 2005)
Bacteriophage phi29 mature particle	16	2006	(Xiang *et al.*, 2006)
Bacteriophage phi6 nucleocapsid	7.5	2006	(Huiskonen *et al.*, 2006)
Bacteriophage phi6 virion	18	2007	(Jäälinoja *et al.*, 2007)
Bacteriophage phi8 core	8.7	2007	(Jäälinoja *et al.*, 2007)
Bacteriophage phi8 virion	21	2007	(Jäälinoja *et al.*, 2007)
Bacteriophage phiKZ	28	2005	(Fokine *et al.*, 2005)
Bacteriophage phiKZ capsid	18	2005	(Fokine *et al.*, 2005)
Bacteriophage PRD1	20	1995	(Butcher *et al.*, 1995)

Virus Name	Resolution (Å)	Year	Reference
Blue tongue virus	23	1997	(Grimes *et al.*, 1997)
Bovine papilloma virus	25	1991	(Baker *et al.*, 1991)
Bovine papilloma virus type 1	9	1997	(Trus *et al.*, 1997)
Broadhaven virus	23	1997	(Schoehn *et al.*, 1997)
Brome mosaic virus (native form)	18	1999	(Krol *et al.*, 1999)
Cauliflower mosaic virus	23	1989	(Kovgan & Zhdanov, 1989)
Cowpea chlorotic mottle virus (native form)	24	1995	(Speir *et al.*, 1995)
Cowpea chlorotic mottle virus (Swollen Form)	22	1995	(Speir *et al.*, 1995)
Cucumber mosaic virus	12	2002	(Bowman *et al.*, 2002)
Cytoplasmic polyhedrosis virus	9	2004	(Booth *et al.*, 2004)
Dengue virus (immature)	16	2003	(Zhang *et al.*, 2003)
Echovirus type 12	18	2004	(Bhella *et al.*, 2004)
Equine herpesvirus 1	45	1990	(Baker *et al.*, 1990)
Fish nodavirus	23	2002	(Tang *et al.*, 2002)
Flock house virus	22	1994	(Cheng *et al.*, 1994)
Hepatitus B	9.1	1997	(Conway *et al.*, 1997)
Hepatitus B	7.4	1997	(Böttcher *et al.*, 1997b)
Herpes Simplex virus A capsid	26	1994	(Zhou *et al.*, 1994)
Herpes Simplex virus B capsid	8.5	2000	(Zhou *et al.*, 2000)
Human Adenovirus Type 5	10.3	2005	(Fabry *et al.*, 2005)
Human Coxsackievirus A21 Complex with ICAM-1 Kilifi	8	2005	(Fabry *et al.*, 2005)
Human papilloma virus	25	1991	(Baker *et al.*, 1991)
Human papilloma virus	60	1970	(Crowther *et al.*, 1970a)
Human parvovirus B19 (B19_Globoside complex)	26	1996	(Chipman *et al.*, 1996)
Human parvovirus B19 (Expressed)	26	1996	(Chipman *et al.*, 1996)
Human rhinovirus	25	1996	(Hewat & Blaas, 1996)
Infectious Bursal Disease Virus	20	1997	(Böttcher *et al.*, 1997a)
Infectious Bursal Disease Virus	10.4	2005	(Saugar *et al.*, 2005)
Infectious Bursal Disease Virus Capsid	12	2005	(Saugar *et al.*, 2005)
Infectious Bursal Disease Virus Subviral Particle	7.2	2005	(Saugar *et al.*, 2005)
Kelp Fly Virus	15	2005	(Hartley *et al.*, 2005)
Nudaurelia Omega Capensis Virus (Capsid Form)	25	2000	(Canady *et al.*, 2000)
Nudaurelia Omega Capensis Virus (Native Capsid Form)	21	2000	(Canady *et al.*, 2000)
Nudaurelia Omega Capensis Virus (Procapsid Form)	28	2000	(Canady *et al.*, 2000)
Pariacoto Virus (Native Form)	21	2001	(Tang *et al.*, 2001)
PM2 Virion	8.4	2004	(Huiskonen *et al.*, 2004)
Poliovirus 135S Particle	8.7	2005	(Bubeck *et al.*, 2005)
Reovirus	30	1991	(Metcalf *et al.*, 1991)

Virus Name	Resolution (Å)	Year	Reference
Rice Dwarf Virus	6.8	2001	(Zhou *et al.*, 2001)
Rice yellow mottle virus	25	2000	(Opalka *et al.*, 2000)
Semliki Forest Virus	35	1986	(Vogel *et al.*, 1986)
Semliki Forest Virus	9	2000	(Mancini *et al.*, 2000)
Simian Virus 40	38	1988	(Baker *et al.*, 1988)
Simian Virus 40	26	1989	(Baker *et al.*, 1989)
Sindbis Virus	35	1987	(Fuller, 1987)
Sindbis Virus	9	2006	(Mukhopadhyay *et al.*, 2006)
Southern Bean Mosaic Virus	25	2000	(Opalka *et al.*, 2000)
Sulfolobus turretted icosahedral virus	27	2004	(Rice *et al.*, 2004)
Tomato Bushy stunt	23	1971	(Crowther, 1971)
Turnip yellow mosaic virus	15	1996	(Böttcher & Crowther, 1996)
West Nile virus (immature)	24	2007	(Zhang *et al.*, 2007)
Western equine encephalitis virus	30	2002	(Mukhopadhyay *et al.*, 2002)
Yellow fever virus (immature)	25	2003	(Zhang *et al.*, 2003)

Naturally, the methods have improved over the years to produce 3D reconstructions of viruses from the resolution in the 1970s of 60 Å and 23 Å for HPV and tomato bushy stunt virus respectively (Crowther, 1971; Crowther *et al.*, 1970a) up to the recent 6.8 Å resolution obtained for a rice dwarf virus reconstruction (Zhou *et al.*, 2001). This trend of improvement of resolution over time is illustrated in Figure 1.8.

A number of developments have led to the improvement of resolution in EM reconstructions. Initially negative stain was used to prepare the sample for imaging. This, however, leads to distortion of the virus image due to collapse of the viral shell and uneven staining of the particle. Cryo-EM ensures that the particle remains in its native state and reduces radiation damage of the sample (Dubochet *et al.*, 1985). The use of the highly coherent field emission gun (FEG) also improves the quality of the images by reducing the falloff of the contrast transfer function (CTF), which limits the resolution, so that more information is obtainable at high resolutions (Mancini *et al.*, 1997). In addition to these improvements of data collection methods, the collection of more particles, and the improvement of reconstruction methods has also improved the resolution of recent reconstructions (Figure 1.8). The resolution often also depends on the number of images that are excluded from the reconstruction as improved data lead to improved reconstructions. Ultimately,

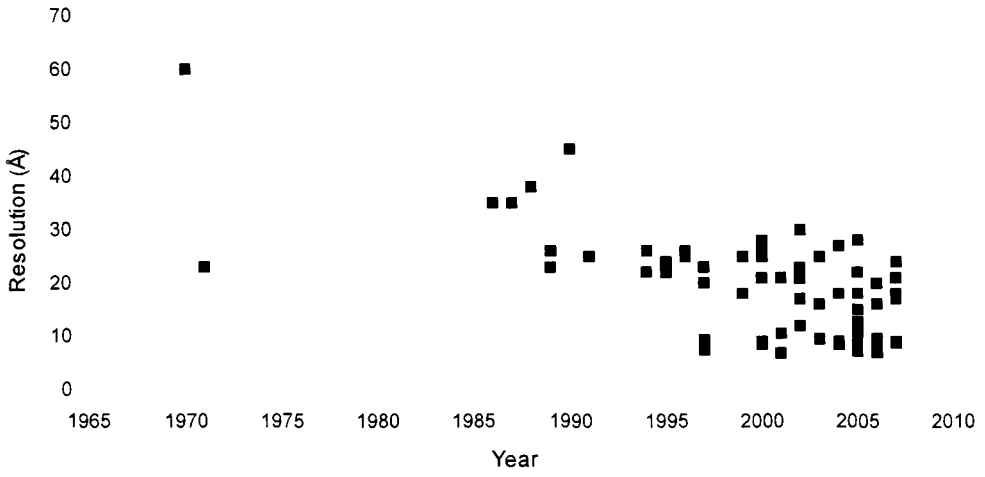

Figure 1.8: *Improvement of resolution of cryo-EM icosahedral reconstructions over the years, including the original negative stain reconstructions by Crowther et al. (1970a, 1971) for comparison.*

the quality of the reconstruction depends on accuracy and precision in the alignment of the particles (Mancini *et al.*, 1997).

The first structures to be solved at subnanometre resolutions were the independent reconstructions of hepatitis B virus capsid to 7.4 Å (Böttcher *et al.*, 1997b) and 9.1 Å (Conway *et al.*, 1997). Böttcher *et al.* (1997b) refined separate models for each of the 34 micrographs collected using the cross common lines method to compare the 6 650 particles to projections. This was used to determine their orientations and phase origins which were used to produced models from the images. The separate models, with different CTFs are then merged into one CTF-corrected model which is projected to iteratively refine the orientations (Böttcher *et al.*, 1997b).

Conway *et al.* (1997) use the polar Fourier transform (PFT) method for determining the orientations of the particles. 2 040 particles were used from two micrographs of the the same sample at different values of underfocus producing a focal pair of images that were combined, with phase reversals to correct for the CTF (Conway *et al.*, 1997). Whereas the Böttcher *et al.* (1997b) final reconstruction used 6 384 images, the Conway *et al.* (1997) refinement excluded over 1 000 images for a final reconstruction from 600 particles. Although an increase in the number of particles increases the SNR and therefore the reconstruction, the use of many particles that exhibit heterogeneity or that have incorrectly determined orientations leads to a poorer reconstruction. The quality of a

reconstruction therefore depends on the fraction of particles excluded as well as the number of particles used for the final map (Mancini *et al.*, 1997).

Using their new Multi-Path Simulated Annealing method, a Monte-Carlo type of optimisation algorithm, to align the particles and determine their orientation, Liu *et al.* (2007) were able to produce a 9.6 Å model of the rice dwarf virus from 62 particles and a 7.9 Å resolution model from 284 particles. These were the best particles selected from a total set of 4 865 particles. Randomly selecting 284 particles from the original data set leads to a poorer resolution of 10.6 Å. Therefore, although high quality reconstructions are possible from few images, it is still necessary to collect many images as the advantage is due to the selection of the best particles (Liu *et al.*, 2007).

1.2.4. Icosahedral program conventions

Single particle reconstruction programs use different conventions for their Euler angles, standard orientations of an icosahedral object and asymmetric unit for the icosahedron. It is necessary to understand the nuances of the programs if they are to be used in combination as the data must be transformed from one format to another.

a) Euler angles

A rotation in three dimensional space may be described by three parameters. Euler angles are a set of three angles that are used to describe the rotation of a three dimensional body by applying three rotational matrices to the positional coordinates of the object. In the SPIDER (System for Processing Image Data from Electron microscopy and Related fields) (Frank *et al.*, 1996) and the MRC (Medical Research Council) software packages (Crowther *et al.*, 1996), the *ZYZ* convention is used whereby the three angles refer to rotation about the *z*-axis, followed by rotation about the *y*-axis and then another rotation about the *z*-axis (Baldwin & Penczek, 2006).

In SPIDER, the rotation is described by the notation (φ, θ, ψ), representing the rotation matrices:

$$\begin{bmatrix} \cos(\psi) & \sin(\psi) & 0 \\ -\sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \begin{bmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The object is first rotated by φ around the z-axis, then θ around the y-axis, followed by ψ around the z-axis again, with all angles describing clockwise notation.

While the MRC programs also follow the *ZYZ* convention, they use the convention (φ, θ, Ω), with $\varphi_{Spider} = -\varphi_{MRC}$, because the rotations are in opposite directions (Baldwin & Penczek, 2006).

However, EMAN (Ludtke *et al.*, 1999) uses a *ZXZ* convention with the rotational matrices:

$$\begin{bmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(alt) & \sin(alt) \\ 0 & -\sin(alt) & \cos(alt) \end{bmatrix} \begin{bmatrix} \cos(az) & \sin(az) & 0 \\ -\sin(az) & \cos(az) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The conversion can be made between EMAN and SPIDER using the following equations to convert the angles (Baldwin & Penczek, 2006):

$$\begin{aligned} az &\equiv \phi_{SPIDER} + \frac{\pi}{2}, \\ alt &\equiv \theta, \\ \phi_{EMAN} &\equiv \psi - \frac{\pi}{2}, \end{aligned}$$

Unless otherwise stated, the Euler angles referred to in this thesis are in the SPIDER convention.

b) Icosahedral conventions

There are different standard orientations for the icosahedra in the various programs. This is particularly important for the MRC icosahedral programs where the orientations that the program determines are relative to its standard orientation.

In the MRC software package (and those from Purdue University and European Molecular Biology Laboratory; EMBL) the icosahedron is orientated so that three of the two-fold axes are aligned with the x-, y- and z-axes such that one three-fold view is ($\theta = 69.09°$, $\psi = 0°$) (Baker *et al.*, 1999). The default orientation in both EMAN and SPIDER is for a five-fold axis to be aligned with the z-axis. In SPIDER this is such that when the structure is viewed from the positive z-axis, an edge points along the positive x-axis (Lebarron, 2005). EMAN rotates the structure 180° (or 36°) about the z-axis. (See Figure 1.9)

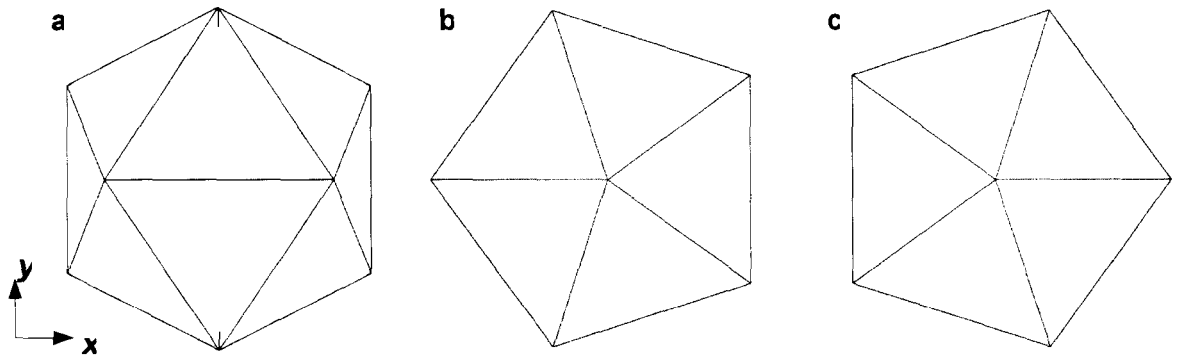

Figure 1.9: *Standard icosahedral orientations for* ***a)*** *MRC,* ***b)*** *SPIDER and* ***c)*** *EMAN reconstruction programs viewed down the z-axis. (Images made with Chimera; Pettersen et al., 2004.)*

A different orientation can be used in SPIDER (Frank *et al.*, 1996) by converting the icosahedral symmetry file that is provided for the reconstruction, to the correct orientation. The symmetry file contains a list of trios of Euler angles that, when used to rotate a symmetrical object, leave it unchanged. EMAN 1.8 (Ludtke *et al.*, 1999) contains functions to rotate objects between different icosahedral orientations.

c) Asymmetric units

It is necessary to reconstruct only one asymmetric unit of the molecule because the rest can be generated by symmetry. In the MRC programs, this is bounded by two five-fold axes at (90°, ± 31.717°), and the adjacent three-fold axis (69.09°, 0°) (Fuller *et al.*, 1996).

Chapter 2. Data Collection

2.1. Introduction

A purified and concentrated sample is required for EM data collection. The sample should be concentrated enough such that each micrograph recorded contains a respectable number of virus particles yet not too concentrated that the virus particles overlap one another. Typically $10^{11} - 10^{12}$ virus particles per microlitre is ideal (Thuman-Commike & Chiu, 2000). The sample is loaded onto a carbon-coated copper grid in order to be examined with an electron microscope.

One of the main considerations in data collection is the radiation damage caused by the high energy electron beam through the sample, which reduces the overall resolution of the images collected. Additionally, because the electron beam needs to pass through a vacuum to prevent the electrons being scattered by air molecules, the sample needs to be stable under low pressure (Auer, 2000). The various preparation and data collection techniques are designed in order enhance contrast of the data and reduce this radiation damage.

Initial contrast enhancement methods involved shadowing the sample with heavy-metal atoms or positive staining with metal atoms that react chemically with the specimen in order to increase the electron scattering of the specimen (Crowther, 2004). The two main methods currently used to prepare biological samples for data collection for 3D reconstruction, are negative staining and vitrification. Variations of these methods include cryo-negative staining, the carbon-sandwich technique (the stained specimen is embedded between two carbon films), and the addition of glycerol or glucose to reduce the artefacts due to dehydration and incomplete stain embedding (Ohi *et al.*, 2004).

The term, "negative stain", is derived from the fact that the contrast is generated by the stain, which surrounds the biological material, leaving the virus particles, which have very low scattering power, as holes in the electron dense stain (Hall, 1955). While this method allows surface features to be identified, information about the internal molecular structure is limited by the penetration of the stain into the sample cavities (Crowther, 2004). Another limitation of the method is lower resolutions, limited to approximately 20 Å, due to radiation damage from the electron beam and the presence of microcrystals of heavy atoms that form upon drying (Ohi *et al.*, 2004). The particles also tend to be distorted and flattened by the drying of the sample. Negative staining of the sample with 1% phosphotungstic acid was immediately recognised as an good method for determining the

ultrastructure of viruses (Brenner & Horne, 1959). Other heavy-metal solutions, of which uranyl acetate is the most common (Ohi *et al.*, 2004), are also used to generate contrast in a sample, which is then air dried on the grid.

Cryo-EM was developed by Dubochet and his colleagues at the EMBL, Heidelberg, Germany (Dubochet *et al.*, 1985). With cryo-EM, radiation damage of the sample is reduced by collecting the data at liquid nitrogen temperatures (-170 °C). The low temperature traps the free radicals formed by the inelastic interactions of the electrons with sample atoms to prevent the spread of the radiation damage (Auer, 2000). In addition, the biological sample is maintained closer to its native state since it is embedded in vitreous ice. The biological sample itself, rather than the stain, is imaged allowing for visualisation of internal structure. Unfortunately the amplitude contrast of the collected data is very small when the microscope is in focus as the ice and biological material allow most of the electron beam to pass through. Therefore the data are collected slightly under focus in order to create phase contrast as described in Chapter 3 to aid in the alignment of the images.

In addition to the cryo-protection from radiation, cryo-EM is also advantageous over negative staining as the frozen hydrated species is not distorted by dehydration as with negative staining (Saibil, 2000). The higher contrast due to the negative staining may be helpful in identifying heterogeneity of the sample (Doan *et al.*, 2003; Ohi *et al.*, 2004), however, the cryo data in general result in reconstructions with higher resolution.

It is necessary for the samples to be frozen rapidly, using efficient cryogens such as liquid ethane or propane, to form vitreous rather than crystalline ice. A thin layer of the sample solution ensures that the cooling is rapid enough, the individual particles are not overlapping and the ice is not too thick for the electrons to penetrate (Baker *et al.*, 1999). Excessive drying, however, thins the sample and can cause the particles to migrate to the edge of the film, where it tends to be thicker, hence increasing the particle concentration and possibly altering the specimen's structure (Baker *et al.*, 1999). The sample preparation is therefore carried out under high humidity conditions to prevent dehydration, with the thickness of the sample controlled by blotting. Additionally liquid nitrogen or liquid helium temperatures must be maintained throughout the data collection process to avoid devitrification (Baker *et al.*, 1999).

The low-dose data collection method used to reduce radiation damage involves a three step process. First, a suitable area for recording is found using low magnification (< 2 000 × to 3 000 ×) with low irradiation levels (< 0.05 $e^{-}.Å^{-2}.s^{-1}$) to prevent the sample from being damaged. Second, focussing and astigmatism correction is carried out under high magnification (> 100 000 ×) on an area adjacent to that being recorded to prevent radiation damage. Finally, intermediate magnification (25 000 × to 50 000 ×) and an electron dose of between 5 and 20 $e^{-}.Å^{-2}$ are used for recording images. Care must be taken to ensure that the sample area to be captured receives minimal electron beam exposure and thus well planned collection strategies are used so that the area photographed does not overlap with areas previously used to focus the beam (Baker *et al.*, 1999).

Once micrographs of the sample have been collected, the next critical step is particle picking. The particles are picked from the micrographs into square boxes either manually or by using automatic picking programs, *e.g.*, BOXER from the EMAN suite (Ludtke *et al.*, 1999). An optimal box size is used such that it is large enough to prevent edge effects from further processing affecting the central particle and small enough to prevent long processing times.

2.2. Methods

2.2.1. Virus sample preparation

A virus sample was provided by Dr Keizo Nagasaki's research laboratory (National Research Institute of Fisheries and Environment of Inland Sea, Hiroshima, Japan). The virus was isolated from infected *H. circularisquama* using the method described by Tomaru *et al.* (2004). Briefly, 450 ml of exponentially growing *H. circularisquama* strain, HCLG-1, was inoculated with 3 ml of HcRNAV109 producing approximately 107 infectious titre.ml^{-1}. The cells were lysed and filtered sequentially through 8.0, 0.8 and 0.2 µm filters to remove cell debris. Polyethylene glycol was added to the filtrates to obtain a 10% (w/v) final concentration and stored in the dark overnight at 4 °C. The suspension was centrifuged at 57 000 × *g* for 90 min. The viral pellet was then washed with phosphate buffer (10 mM Na_2HPO_4 and 10 mM KH_2PO_4 in distilled water) and centrifuged again at 21 7000 × *g* for 4 h. The virus particles thus collected were then resuspended in 500 µl 10 mM phosphate buffer.

The concentration necessary for cryo-EM is approximately 10 to 100 fold more than that of negative staining. It was therefore necessary to concentrate the sample using a 100 000 kDa cut off spin filter from PAL centrifuged at 4 000 × g for 2 – 3 min.

2.2.2. Negative stain electron microscopy

Negative stain was first used to view the particles and ascertain the concentration. The sample was prepared by applying 3 µl of virus sample to a glow-discharged carbon-coated grid and allowing it to be absorbed for 30 s before blotting. The grid was then successively floated on two drops of water and three droplets of 2% uranyl acetate placed on Parafilm™, blotting with filter paper between washings. The air-dried grid was then viewed under the Leo 912 TEM.

Data were recorded using low dose methods on Kodak electron image film SO-163 on a Leo 912 TEM operating at 120 kV and a nominal magnification of 50 000 ×. The film was developed in Kodak Professional D19, fixed with Ilford Rapid Fixer and digitised using a Leafscan 45 by Ilford scanning at 10 µm.pixel^{-1} for micrographs with a sampling rate of 2 Å.pixel^{-1}.

2.2.3. Cryo-electron microscopy

Cryo-EM facilities at the MRC Laboratory of Molecular Biology (LMB), Cambridge, UK were used for collecting the cryo-EM data. The sample was prepared for cryo-EM in a humidity-controlled fume-hood in order to prevent excessive evaporation. 3 µl of sample were applied to 300 mesh copper R2/2 Quantifoil® grids that had been air glow-discharged for 30 seconds to enhance the spreading of the specimen. The grids, suspended by tweezers in a cryo-plunger, were blotted with Whatman no. 1 filter paper for 1 to 3 seconds to form a very thin aqueous film. Before the liquid film could dry, the grids were immediately submerged in ethane slush cooled by liquid nitrogen by releasing the plunger (Bellare *et al.*, 1988). This resulted in the virus particles being embedded in a thin layer of vitrified ice over the carbon holes. Liquid nitrogen temperatures were maintained throughout the subsequent data collection steps by transferring the sample grid to a side-entry Gatan 626 cold stage to be loaded into the electron microscope.

An FEI Tecnai F30 FEG microscope operating at 300 kV and a defocus range of 1.6 to 5 µm was used to collect images of HcRNAV109 under low-dose (about 10 electrons.Å^{-2}) conditions at

55 299.5 × magnification (nominal magnification of 59 000 ×). The images were recorded on Kodak SO163 film developed in D19. The micrographs were visually examined to discard those containing few or no particles and too much ice contamination. 102 of the 108 micrographs recorded were scanned at 10 μm.pixel^{-1} using a Leafscan 45 by Ilford to give sampling of 1.808 Å.pixel^{-1}.

2.2.4. Computer hardware and software

Image processing and 3D reconstructions were performed on four Intel servers with two 2.8 GHz Xeon processors and 2 GB RAM running Redhat Enterprise Linux. Initial image processing was performed using SPIDER 13 and 15 (Frank *et al.*, 1996) and EMAN 1.7 (Ludtke *et al.*, 1999) single particle reconstruction software. WEB (Frank *et al.*, 1996) and v2 (Ludtke *et al.*, 1999) were used for viewing images. *Refine* and *fftrans*, from the MRC software package (Crowther *et al.*, 1996), were used to determine the initial orientations of the particles while SPIDER was used for refinement of the orientations. Chimera (Pettersen *et al.*, 2004) was used for visualising the 3D models and creating images.

2.2.5. Particle picking

The micrographs collected by both negative stain and cryo-EM were converted from RAW format (from the scanner) to SPIDER format, interpolated to half the original size, and then copied to MRC format using SPIDER (Frank *et al.*, 1996). According to the Shannon-Nyquist theorem (Shannon, 1949), which requires a sampling at least twice as fine as the resolution desired, this allows for a theoretical resolution of up to 8 Å for the negative stain data and 7.2 Å for the cryo data, although oversampling is desirable and a poorer resolution is expected. The power spectra of the micrographs were computed by SPIDER (Frank *et al.*, 1996) and examined with v2 (Ludtke *et al.*, 1999), in order to discard those which were astigmatic or showed drift by possession of nonisotopic or incomplete Thon rings in the power spectra. Boxer, an EMAN program (Ludtke *et al.*, 1999) for selecting particles from micrographs, was used to automatically pick virus particles from each micrograph with box sizes approximately double the diameter of the particle (200 × 200 pixels). The micrographs and picked particles were then manually examined to refine the particle selection. These particles were converted into SPIDER format for further processing.

2.3. Results

2.3.1. Negative stain

Thirty three negative stain micrographs with original sampling of 2 Å.pixel^{-1} showed virus particles with clear contrast (example of one micrograph in Figure 2.1).

Figure 2.1: A representative electron micrograph of a negative stain preparation of HcRNAV109 collected on a Leo 912 TEM at 120 kV.

Interpolation reduced the sampling to 4 Å.pixel^{-1} which allows for the 3 511 particles with diameters of 344 – 400 Å (86 - 100 pixels) to be picked in 200 × 200 pixel boxes. Inspection of the picked particles revealed the presence of two distinct types of particles. Particles either had a solid, white centre or a darker centre (Figure 2.2).

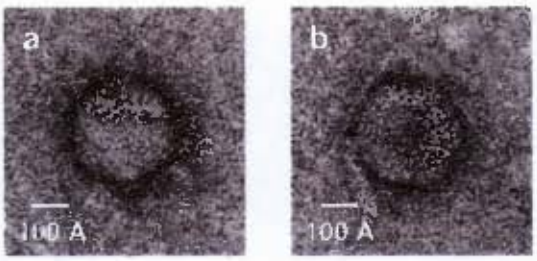

Figure 2.2: Representative negative stain HcRNAV109 particles showing ***a)*** *a solid, compact particle and* ***b)*** *a particle with dark centre*

2.3.2. Cryo-electron microscopy

One hundred and eight cryo-EM micrographs, approximately 5 900 × 9 200 pixels, with a sampling of 1.808 Å.pixel^{-1} were recorded. After interpolation, this resulted in images with 3.616 Å.pixel^{-1} sampling. A sample micrograph is shown in Figure 2.3 where the lighter background is the carbon film and the darker background is the vitreous ice in the hole of the carbon film.

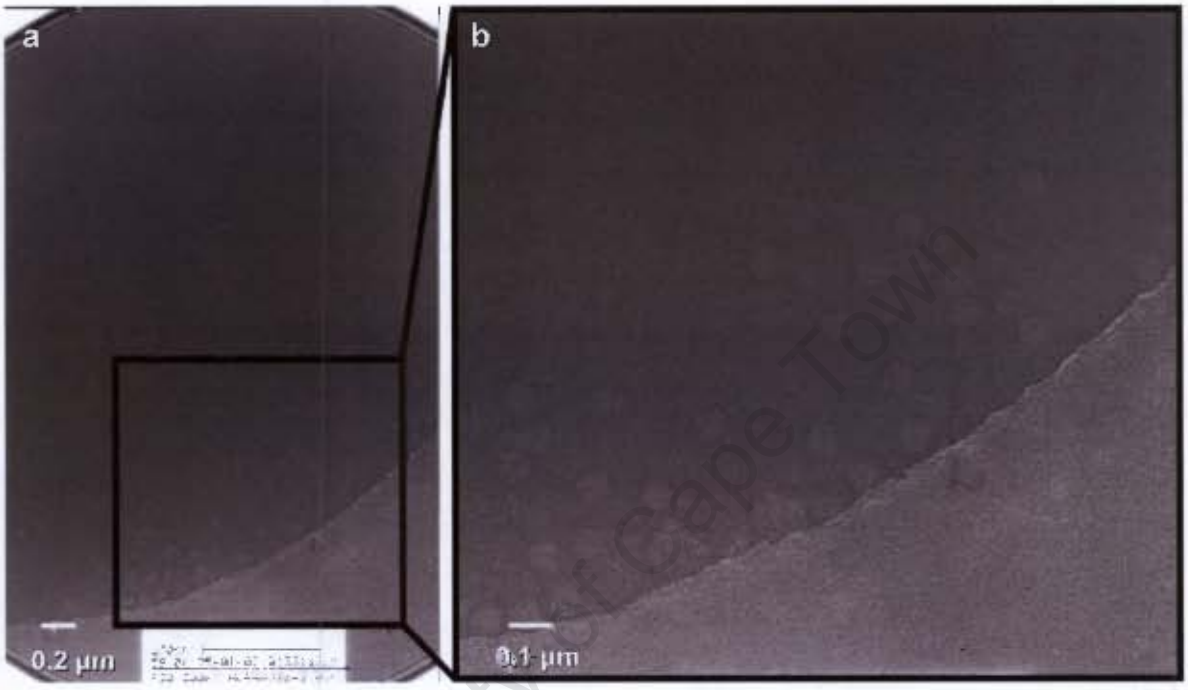

Figure 2.3: ***a)*** *A representative micrograph of HcRNAV109 obtained by cryo-EM;* ***b)*** *A close up of some of the virus particles embedded in vitreous ice in the hole of the carbon film.*

In total 2 593 homogeneous HcRNAV109 particles, with diameters between 318 and 340 Å (88 – 94 pixels), that were situated in the vitreous ice and were clearly distinct from the neighbouring particles were picked from 102 micrographs. The effect of defocus on the images can be seen in Figure 2.4 where particles picked from two different defocus groups are shown.

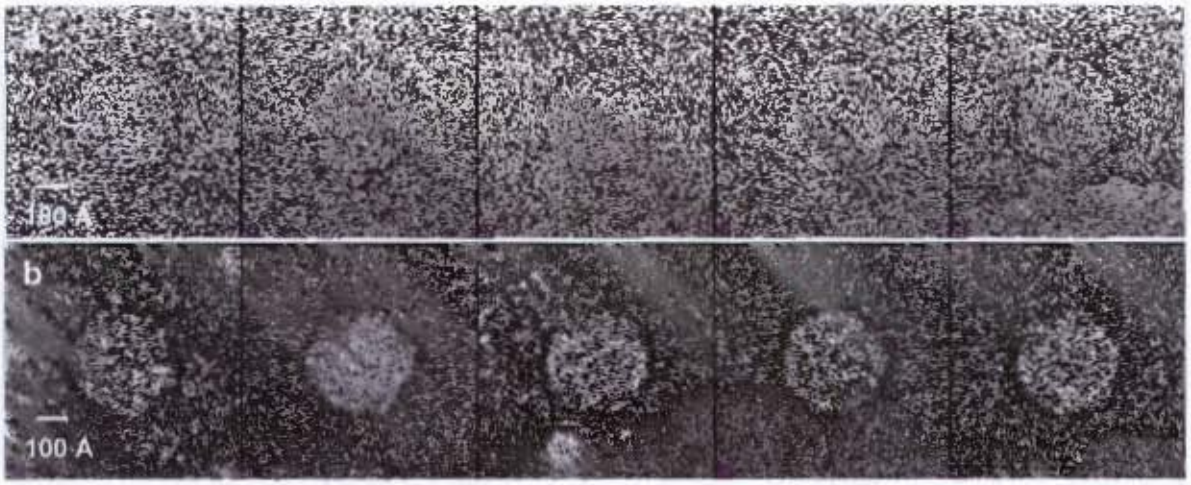

Figure 2.4: *Five picked particles from each of two defocus groups: **a)** defocus group 1 at 1.6 μm defocus and **b)** defocus group 19 at 4.8 μm defocus.*

2.4. Discussion

The range in diameters that is quoted is not due to variation in the sample but rather to the deviation from a sphere of an icosahedral particle. The larger diameters (by 26 – 60 Å) of the negative stain particles is consistent with the reported flattening effect of negative staining on samples (Boisset *et al.*, 1995). The particle distortion is also most likely the cause behind the heterogeneity of the negative stain data as the particles collapse onto the solid carbon film. While the negative stain data do show much greater contrast than the cryo data, the resolution is poorer.

Because of the low concentration of virus, the particles are sparsely distributed in the cryo micrographs. Due to lack of material, further concentration and data collection were not possible. However, the high symmetry of the icosahedral particles allows for high resolution reconstruction from fewer particles.

Chapter 3. Preprocessing

3.1. Introduction

There are a number of processing steps which are necessary before reconstruction can be undertaken. First, particles must be processed to remove unnecessary data that would negatively impact the reconstruction. Second, the defocus values of the different micrographs must be determined. The reconstruction is performed in separate defocus groups and it is therefore necessary to determine the contrast transfer function (CTF) of each micrograph in order to combine the particles from micrographs with similar defocus values so as to increase the number of particles in each defocus group. The defocus values are also used later on in the reconstruction in order to reverse the effect of the CTF on the model.

Amplitude contrast is created when electrons are lost due to scattering by the specimen, outside the objective aperture. In the thin specimens used in TEM, only a very small fraction of the electron wave is scattered producing very little amplitude contrast. Phase contrast, however, is produced when the phase of the electron wave passing through the specimen is shifted depending on the spherical aberration of the microscope and the amount of underfocus at which the image is recorded (Erickson & Klug, 1971).

The image that is captured by the imaging device in the electron microscope is then not simply the projection of the object, but is convoluted with the CTF of the electron microscope which produces resolution-dependent amplitude modulations and phase reversals, called Thon rings (Erickson & Klug, 1971; Thon, 1966). The CTF is due to both the defocus of the microscope, which is used to produce the contrast necessary for alignment, and the spherical aberration inherent in the microscope (Erickson & Klug, 1971). The image can best be deconvoluted in Fourier space since the intensity and phase of the Fourier transform vary with spatial frequency. The power spectrum, which is the square of the Fourier transform, of two electron micrographs taken at different levels of underfocus is shown in Figure 3.1.

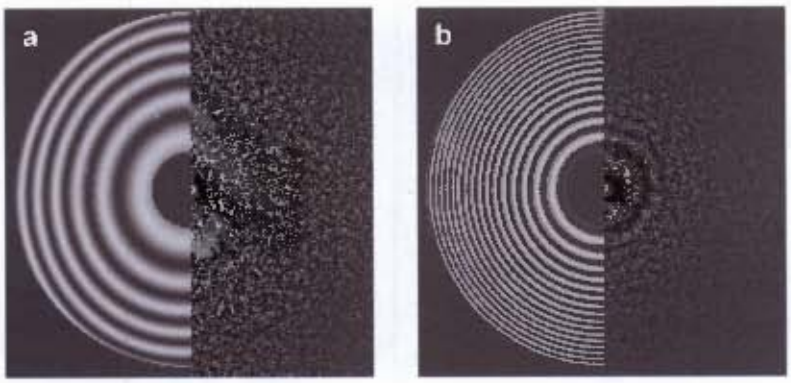

Figure 3.1: *Fourier transforms of two micrographs with defocuses of* ***a)*** *1.6 μm and* ***b)*** *4.8 μm. The average, background-subtracted power spectrum is shown on the right, and the fitted, computed CTF on the left. The power spectra were determined with CTFFIND3 (Mindell & Grigorieff, 2003) and SPIDER (Frank et al., 1996).*

Because the Fourier transform is rotationally invariant, when there is no astigmatism, it may be described mathematically as a function of spatial frequency, s. In Fourier space, the image, $I(s)$, is a combination of the structure-factor intensity, $F^2(s)$, the CTF, the envelope function, $E(s)$, and $N(s)$, the incoherent background noise due to inelastic scattering, ice thickness and noise determined empirically (Ludtke *et al.*, 1999; Zhu *et al.*, 1997):

(1) $I(s)=F^2(s)CTF^2(s)E^2(s)+N^2(s)$

where,

(2) $CTF(s)=-k[\sqrt{(1-Q^2)}\sin(\gamma)+Q\cos(\gamma)]$

(3) $\gamma=-\pi(\frac{1}{2}C_s\lambda^3 s^4-\Delta Z\lambda s^2)$

and k is a scaling constant, Q is the fractional amplitude contrast depending on the electron energy and specimen thickness, C_s is the spherical aberration coefficient of the objective lens, λ is the electron wavelength and ΔZ is the objective lens defocus (with underfocus positive) (Ludtke *et al.*, 2001; Zhou *et al.*, 2003). The CTF is plotted as the thin blue line in Figure 3.2.

The envelope function is due to a combination of instrumental and experimental effects such as spatial and temporal coherence of the electron microscope, specimen movement, recording medium, and scanner optics. It leads to the decay of the Fourier amplitude with increasing spatial frequencies and limits the resolution. Although it has many contributing factors the envelope function may be

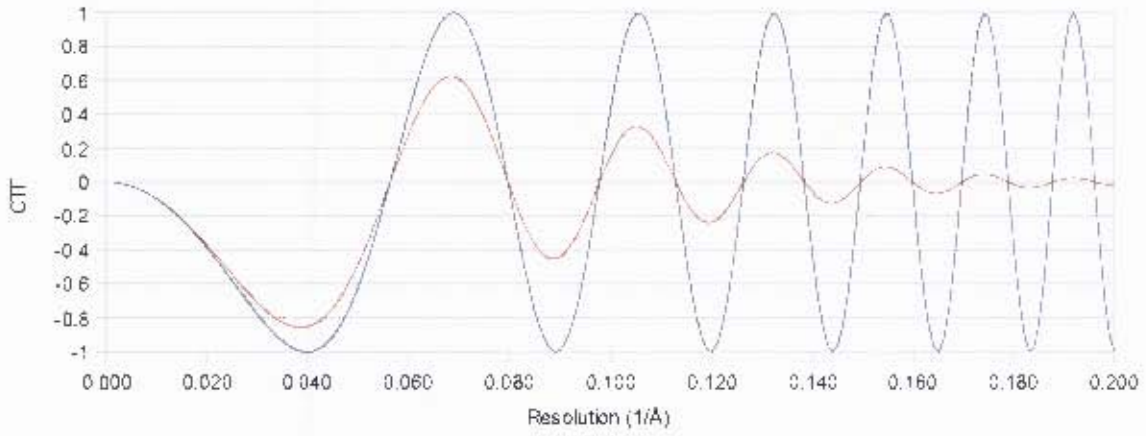

Figure 3.2: *Theoretical contrast transfer function for a 300 kV electron microscope with 2.0 mm spherical aberration at 1.6 μm defocus. (——) represents the CTF with no envelope function; (——) shows the effect of a Gaussian envelope function with B= 100 Å². Generated using a web-based CTF simulation program (Jiang & Chiu, 2002)*

estimated by a single Gaussian envelope function with width, B, estimated experimentally (Saad *et al.*, 2001):

$$E(s)=\exp(-Bs^2) \quad (4)$$

The effect of the envelope function on the CTF can be seen in Figure 3.2.

The amplitude contrast is represented by the cos(γ) term in equation (2) while the phase contrast is represented by the sin(γ) term. As Q takes on values between 0.07 and 0.14 for cryo-specimens and 0.19 to 0.35 for negative stain specimens (Zhou & Chiu, 2003), it is evident from the CTF equation that the phase contrast component will be much more important than the amplitude contrast. A negative (overfocus) or small positive (underfocus) ΔZ would lead to a negative γ and thus opposite signs for the sine and cosine terms of equation (2) so that the value of the CTF is small. The poor low-resolution image contrast that this produces is the reason for recording images significantly underfocus (approximately 1 – 4 μm; Zhou & Chiu, 2003). An increase in defocus leads to an increase in frequency of the CTF.

The only parameter which varies in a given experimental setup is the defocus, ΔZ. Figure 3.3 shows the CTF for a 300 kV electron microscope at two different defocus values. At low resolution, before

the CTF crosses zero, the magnitude of intensity is altered by the CTF, but the phases remain the same. Therefore, for low resolution data, such as negative stain data, which do not extend past the first zero of the CTF, CTF correction is unnecessary. However, at higher resolutions, when $\sin(\gamma)$ goes through zero and changes sign, artefacts are introduced as the contrast of image details corresponding to the frequency are reversed (Erickson & Klug, 1971) and data are lost as the intensity of certain frequencies approaches zero.

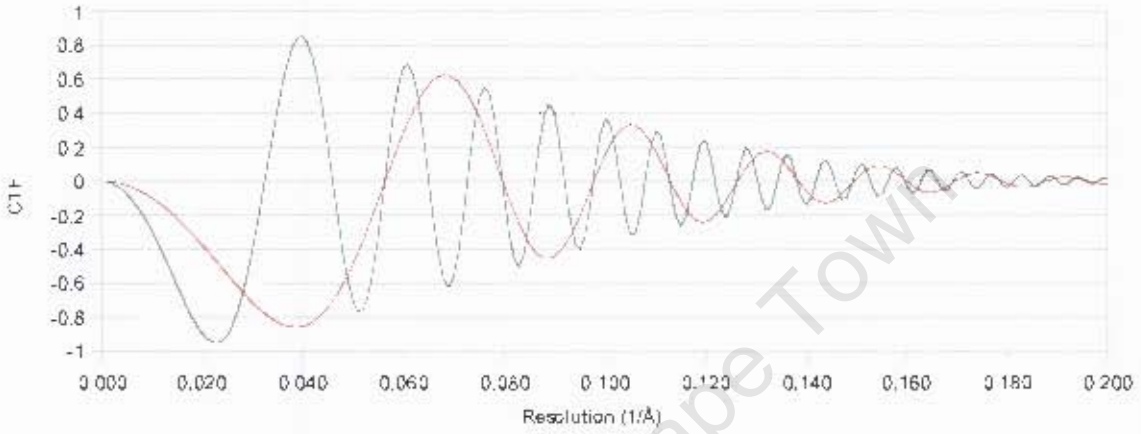

Figure 3.3: *Theoretical contrast transfer function of a 300 kV electron microscope with 2.0 mm spherical aberration at two defocus values: (——) 1.6 µm and (——) 4.8 µm defocus. Generated using a web-based CTF simulation program (Jiang & Chiu, 2002).*

Data lost by the CTF that becomes zero at certain frequencies are compensated for by collecting data at a range of defocus values. This requires the defocus of the data to be estimated by fitting a CTF curve to the averaged Fourier transform of the micrograph or single particles. The transform is clearer if the micrograph contains some carbon, rather than just ice, or a high concentration of protein sample.

Various CTF-correction strategies (elucidated in Chapter 5) are then applied to correct for the distortion of the data due to the CTF. Some of these involve processing the data in separate defocus groups whereby particles with similar defocus values are grouped together.

A number of preprocessing steps, common to all reconstruction methods, must be performed on the images before reconstruction is undertaken in order to optimise future steps: filtering, centring, masking and normalising.

In order to eliminate noise from the image, data of a higher frequency than the expected resolution should be removed. Similarly low frequency information, which is due to stain variation, should also be removed. This is done by applying Gaussian Fourier filters to remove the frequencies without introducing artefacts caused by sharp Fourier filters.

The particles are initially centred, based on their average centre of gravity, with integral shifts in order to avoid interpolation errors. Interpolation errors occur when an image is rotated or translated such that the position of the old pixels is not the same as that of the new pixels. Each pixel then takes on the average of two or more pixels, decreasing the resolution of the image. This centring, which is later refined during further processing, ensures that particles are not cut off by masking.

A soft Gaussian mask, slightly larger than the particle ensures that adjacent particles do not affect the processing of the particles without causing the images to align to a sharp mask rather than the particles in further reconstruction steps.

In order that the particles contribute equally to the reconstruction, the images are normalised to a mean of one and a standard deviation of zero. This is performed after masking once other particles or dust specks are excluded so that they do not affect the image statistics.

3.2. Methods

3.2.1. CTF determination

The Thon rings are stronger when the image contains some of the carbon support film (Mindell & Grigorieff, 2003). Unfortunately these micrographs contain very little carbon and sparsely spread viruses. In order to strengthen the Thon rings so that the defocus may be determined from them, it was therefore necessary to combine the boxed particles into one image, using the montage command from SPIDER (Frank *et al.*, 1996), from which the defocus was calculated.

The defocus of each interpolated micrograph was determined using CTFFIND3 (Mindell & Grigorieff, 2003) which calculates the average CTF from subdivided square tiles, of which irregular areas may be discarded based on their pixel density variance. The following parameters were used: spherical aberration (C_s) = 2.0 mm; electron beam voltage (*HT*) = 300 kV; amplitude contrast (A) = 7%; magnification (XMAG) = 27 649.75 ×; scanner pixel resolution (Dstep) = 10 μm.pixel^{-1}; tile size (Box) = 200 pixels; resolution range (ResMin, ResMax) = 35.0 – 7.5 Å; defocus search range (dFMin, dFMax, Fstep) = 0 – 50 000 Å in 5 000 Å steps. The background was subtracted from the power spectrum, with the power spectrum fitted to a two-dimensional CTF to determine the defocus and astigmatism. The defocus in two perpendicular directions was determined and averaged to give the overall defocus. Micrographs showing astigmatism in the power spectra were disregarded for future processing.

These defocus values were confirmed by generating power spectra using SPIDER scripts (Frank *et al.*, 1996) with the TF ED command and using CTFmatch (Baxter, 2004) to visually confirm the defocus. The micrographs were grouped into defocus groups with a range of under 1 000 Å and average defocus values for each group were determined. The particles from the micrographs in each defocus group were then combined into separate stacks of images.

3.2.2. Pre-processing

The particles were Fourier filtered with SPIDER to 10 Å resolution using a low pass Gaussian filter of 0.36 and a high pass Gaussian filter larger than the particle diameter (156 pixels, 564 Å). This reduced the noise and stain variations respectively. The images were then centred by integer pixel distances with *proc2d* from EMAN (Ludtke *et al.*, 1999), masked with a Gaussian mask with a standard deviation of 20 pixels to a radius of 188 Å (52 pixels) using the SPIDER command, MA (Frank *et al.*, 1996), and normalised using *proc2d* from EMAN (Ludtke *et al.*, 1999), in their separate defocus groups. The image stacks were again carefully examined in order to discard poorly picked or centred particles.

3.3. Results

3.3.1. CTF determination

The power spectra generated by CTFFIND3 are shown in Figure 3.4.

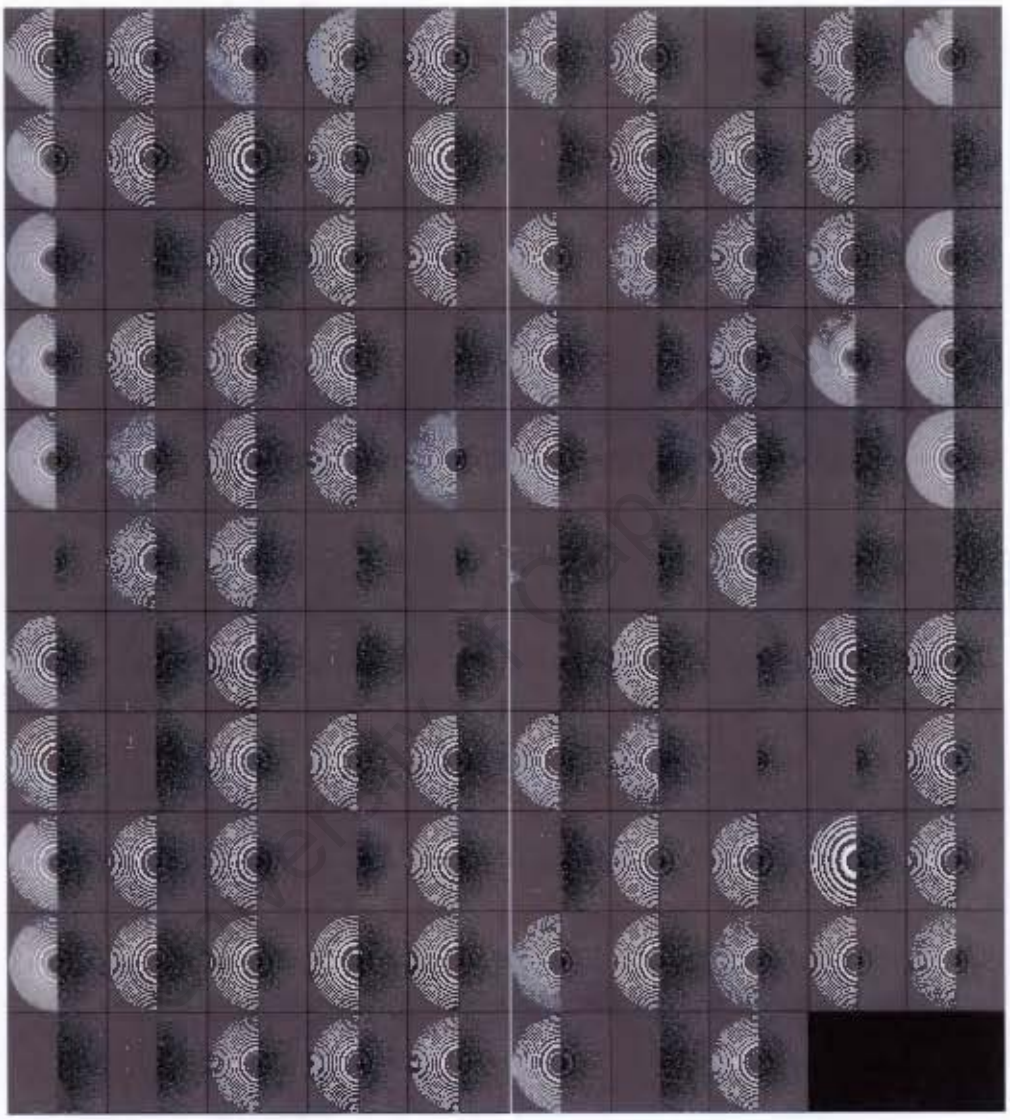

Figure 3.4: *Power spectra of the 108 micrographs recorded. The averaged power spectra are shown on the right and the power spectra computed using the defocus and astigmatism determined by CTFFIND3 (Mindell & Grigorieff, 2003) are shown on the left.*

The defocus values determined from these power spectra are listed in Appendix A. The defocus values were used to split the data into nineteen defocus groups with values of underfocus ranging from 1.6 μm to 4.8μm. This produced image stacks ranging in size from 32 to 374 particles per defocus group. When the average defocus values of each defocus group are calculated and the corresponding CTFs are plotted, there appears to be good coverage of the resolution ranges (Figure 3.5).

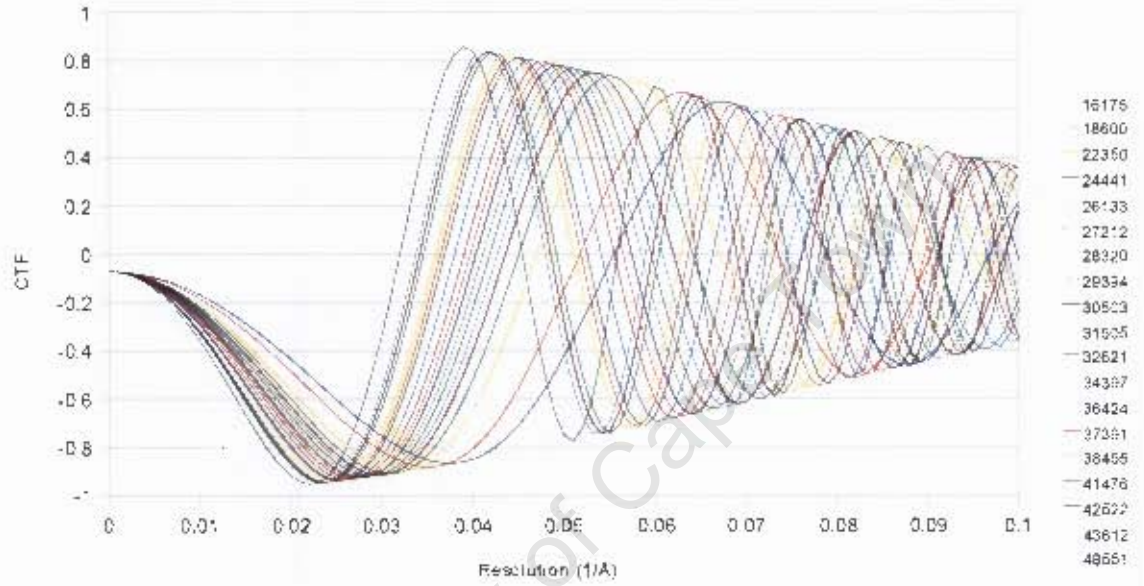

Figure 3.5: *Contrast transfer functions, with envelope functions applied, at the defocus values (in Å) of the nineteen defocus groups used in the reconstruction. CTFs were generated by a CTF simulation program (Jiang & Chiu, 2002).*

However, displaying the CTFs weighted by the number of particles per group (WT = *number of groups* × *number of particles* ÷ *total number of particles*), shows that four groups with defocus values ranging from 3.0 μm to 3.4 μm represent the majority of the particles (Figure 3.6). This leads to data of certain frequencies being more strongly represented to the detriment of other frequencies near the zeros in the frequencies shown in Figure 3.7 where the sum and the sum of the squares of the CTFs are plotted.

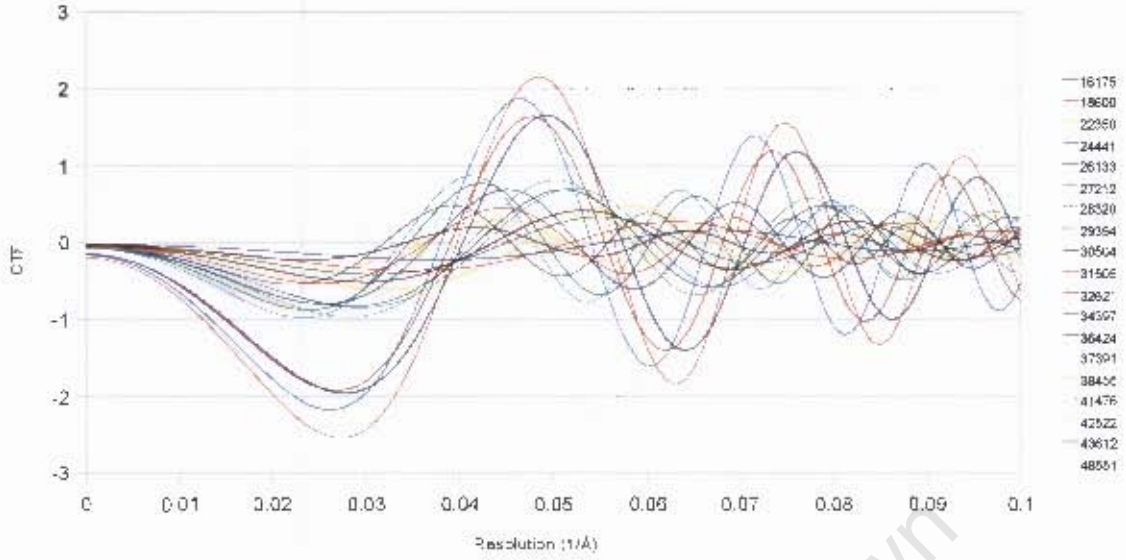

Figure 3.6: *Contrast transfer functions of the nineteen defocus groups weighted by average number of particles per group. CTFs generated by a CTF simulation program (Jiang & Chiu, 2002) with the defocus values in Å listed on the right*

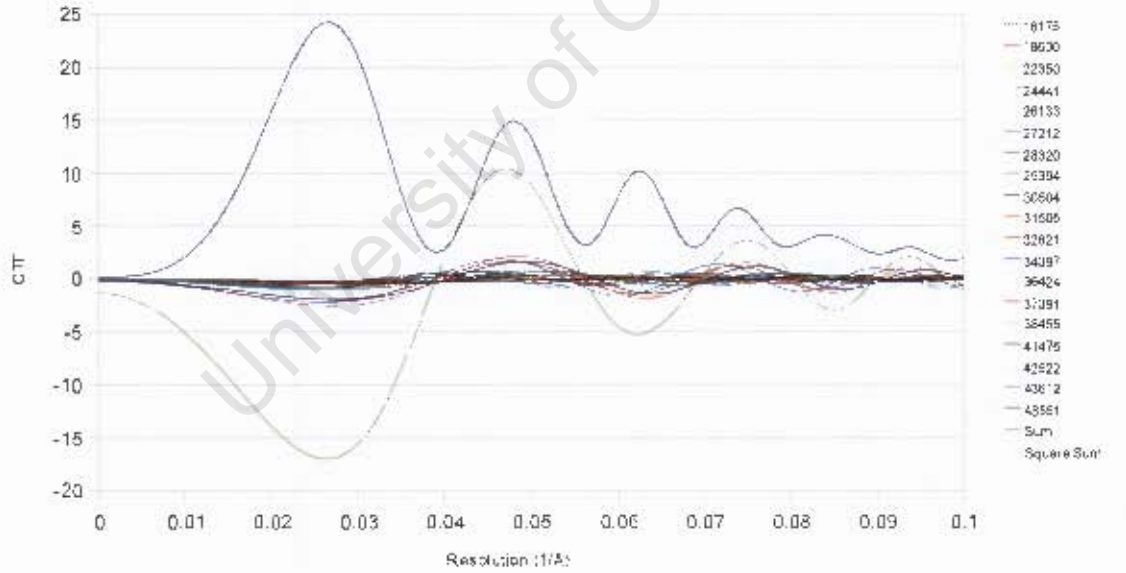

Figure 3.7: *Sum and sum of the squares of the contrast transfer functions over all the defocus groups with CTFs generated by a CTF simulation program (Jiang & Chiu, 2002).*

3.3.2. Preprocessing

A particle at various stages of processing is shown in Figure 3.8. The particle is filtered to remove the high and low frequencies that contain extraneous data, moved to the centre of the image, masked to eliminate the contribution of the surrounding image and normalised to ensure that each image receives the same weighting.

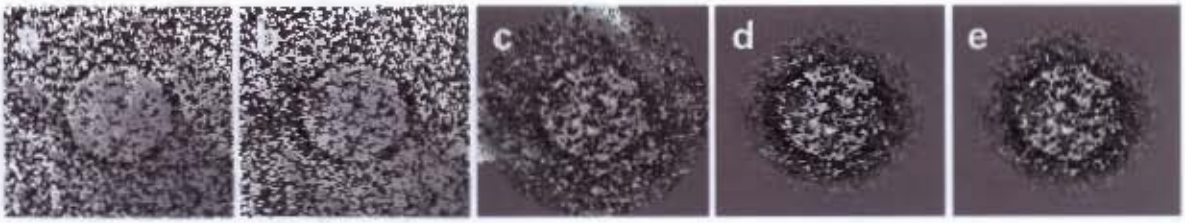

Figure 3.8: *A representative particle,* ***a)*** *before processing,* ***b)*** *Fourier filtered between 10 Å and 364 Å,* ***c)*** *centred,* ***d)*** *masked with a Gaussian filter of radius 52 pixels and standard deviation of 20 and* ***e)*** *normalised.*

3.4. Discussion

The accuracy of the defocus determined for images of single protein particles in ice is generally lower than that of images containing carbon. This is due to the weaker contrast of the images without the strongly scattering carbon, leading to a smaller number of visible Thon rings which causes difficulty in fitting the correct CTF. Varying depths of the particles in ice will also impede accurate defocus determination (Mindell & Grigorieff, 2003).

The consistency between the defocus values predicted by CTFFIND3 (Mindell & Grigorieff, 2003) and visual examination of the plots of CTFs generated by SPIDER's TF ED command (Frank *et al.*, 1996) in comparison to calculated CTFs with CTFMatch (Baxter, 2004) strengthens the reliability of the values, and therefore the defocus grouping. In retrospect, more care should have been taken during data collection to ensure almost equal numbers of particles per defocus group. This is corrected for during refinement by limiting the number of particles used in those defocus groups to only the best 70 in each independent reconstruction (as assessed by correlation scores for alignments to projections of the model).

Chapter 4. Starting Models

4.1. Introduction

An initial starting model is required for refinement of the orientations of the particles. The starting model may be based on structures of related particles or an already determined model of the particle (such as previous EM studies, X-ray crystallography or homology modelling). If there are no previous structural studies then the model must be generated from the data using the common lines methods as described below, or conical tilt methods. Conical tilt involves collecting a pair of images at 0° and at high (60°) tilt to obtain a set of views with known orientations (Radermacher *et al.*, 1987; Saibil, 2000). Another option is to simply use an amorphous model of the appropriate size, which, with iterative refinement, converges on the correct model. Once an initial model is obtained, projection matching can be used to refine the model. In order to determine resolution, ideally two independent starting models should be used.

The principle of common lines depends on the central section theorem whereby the Fourier transform of an image, which is the two-dimensional projection of the 3D object, is a central section of the Fourier transform of the 3D object taken normal to the direction of projection (DeRosier & Klug, 1968; Fuller *et al.*, 1996). Therefore the transforms of the two-dimensional projections of different views of an object may be combined to produce the 3D transform of the object, which in turn can be transformed back into a model of the original 3D object by Fourier inversion (Crowther *et al.*, 1970b).

This, however, requires knowledge about the orientation of the particles relative to each other as well as the symmetry axes, if symmetry is to be used. There are a number of different methods that can be used: real space common lines (van Heel *et al.*, 2000), Fourier common lines (Crowther *et al.*, 1970b) and the polar Fourier transform method (PFT; Baker & Cheng, 1996).

Fourier common lines can be used to determine the orientation of icosahedral particles without a reference model. Since the Fourier transform of two projections both form central sections of the Fourier transform of the 3D object (DeRosier & Klug, 1968), they will intersect forming a common line along which they have the same values. These cross-common lines can be used to determine the relative orientations of the two particles (Crowther, 1971).

More importantly for the creation of a starting model, this same principle can be used to determine the orientation of a single projection relative to the symmetry axis. When a particle is viewed down a symmetry axis, the same symmetry is seen in the projection and the orientation of the particle can therefore be inferred. When the symmetry axis is not along the direction of view, then common lines can be used to determine the orientation. A symmetry operation applied around the symmetry axis (e.g. rotation by 120° around a three-fold axis) to the transform produces a second identical plane which intersects along the common line. Applying the inverse of the symmetry operation (rotation by -120° in this instance) produces another plane with a common line with identical values to the original one forming a pair of common lines. This can be repeated for all the symmetry operations and their inverses. In an icosahedron, the 6 five-fold axes (with two common line pairs each), 10 three-fold axes (one pair of common lines) and 15 two-fold axes (where the two common lines forming the "pair" are co-linear and thus centrosymmetric with phases of 0° or 180°), produce a total of 37 pairs of common lines (Fuller *et al.*, 1996).

The position of the pairs of common lines is known for different orientations of an icosahedron. The phase residual between pairs of common lines is computed over orientations of the asymmetric unit. The angular parameters that give the smallest sum of differences for all the pairs of common lines correspond to the predicted orientation (Crowther, 1971). This works best for orientations near the centre of the asymmetric unit as the common lines are well spread. Along or near a symmetry axis the lines overlap leading to less trustworthy results. The low SNR of cryo-images makes them particularly susceptible to this (Fuller *et al.*, 1996) necessitating the need for the use of class averages.

Due to the different views of particles present in a data set, and sometimes heterogeneity of the sample, it is necessary to classify the images so that class averages may be produced with increased SNR. The first step is a statistical analysis of the image data in order to determine similarity measures between images. Second, these images are grouped into classes based on their similarity (Frank, 2006).

Image similarity is measured by comparing corresponding pixels of aligned images and computing their mean squared difference. Comparing every pixel in order to classify a data set without references is an n^2-dimensional problem (where image size is $n \times n$ pixels), which would be extremely computationally intensive. In practice, pixels do not vary independently, but in

combination with surrounding pixels (Frank, 2006). Multivariate data analysis (MDA), also known as multivariate statistical analysis (MSA), is used to reduce the total interimage variance into mutually orthogonal eigen images that combine the common image features with minimum loss of information. This reduces the dimensionality of the problem as each image is a linear combination of the few eigen images. The eigen images can then be used as the basis for clustering the images into classes (van Heel & Frank, 1981). Principle component analysis (PCA) and correspondence analysis (CA) are the two related techniques used to implement this (Frank, 2006).

While distribution of images in factor space may indicate logical classes in which to divide the images, this is not always the case. Automated or semi-automated techniques require methods to determine the number of classes and membership of the classes. The classification may be *supervised,* when the images are classified based on their similarity to a set of references (as used in 3D projection matching), or *unsupervised*, whereby an intrinsic grouping of the set and their mutual relationships is used (Frank, 2006). In order to avoid reference bias, whereby images are aligned to form a model of the reference that they do not necessarily represent, unsupervised classification is generally used first. There are two commonly used clustering algorithms that are then used to determine the classes: *K*-means and hierarchical ascendant classification (HAC). *K*-means is a non-hierarchical clustering method involving dividing the objects into *K* groups based on their similarity to the centroid (average) of each group, which is recalculated iteratively. Penczek *et al.* (1996) use the observation that it is relatively easy to centre single particle projections and thus the only variable is the rotation of each image around its centre. HAC involves successively merging elements into clusters based on their distances in factor space (Frank, 2006).

The success of the classification depends on the accuracy of the alignment of the particles. The alignment, however, improves with comparison to references, which in the absence of previous structural data, need to be generated from classification. Classification and alignment are therefore iterated to improve both the quality of the classes and the alignment of the particles (Thuman-Commike, 2001).

EMAN (Ludtke *et al.*, 1999) uses a different method for the generation of an initial icosahedral starting model. The combined data set is searched for the particles with the best five-, three- and two-fold symmetries by calculating the dot product of the particle with itself after rotation. Each

group of particles with identified views is then mutually aligned and averaged, producing three characteristic views which are used to build the starting model.

Once an initial model is obtained, the orientations and centring may be refined by iteratively aligning the particles to projections of the reference model and reconstructing a new 3D model from the new class averages of the newly aligned particles as described in Chapter 5.

4.2. Methods

4.2.1. Classification

Classification was carried out on a merged data set of all defocus groups. Particles were rotationally aligned and classified into 30 class averages, using rotational alignment with a *K*-means clustering algorithm (AP CA from SPIDER; Frank *et al.*, 1996), and then centred using reference-free alignment as implemented by the AP SA command (Frank *et al.*, 1996). Although the procedure was iterated, the class averages from the first iteration were used for the generation of the initial model from common lines.

4.2.2. Initial model – common lines

The common lines method implemented in the program *refine*, provided by Tony Crowther (MRC LMB, Cambridge, UK), was used to determine the orientation of the class averages (Crowther *et al.*, 1996). The *refine* program determines the Crowther residual, standard deviation (sigma) and the weighted Crowther residual (weighted by the standard deviation), in order to compare the images with the positions of the predicted pairs of common lines for orientations over the asymmetric unit. Class averages were chosen that showed clear structure and that produced self-consistent results (*i.e.* the same orientations predicted by the standard deviation and the weighted Crowther residual) with good statistics. Models were formed by using the iterative algebraic reconstruction technique to back-project one or more class averages with the SPIDER command, BP RP (see Chapter 5 for a discussion on back-projection; Frank *et al.*, 1996), using the orientations determined by *refine* (Crowther *et al.*, 1996). The most defined model was chosen and subjected to one iteration of projection matching and back-projection using the merged set of cryo-data from all the defocus groups in order to confirm whether or not the model is agreed with the data. This model was then used as the initial model in further reconstructions.

A second starting model was produced by Alan Roseman (MRC LMB, Cambridge, UK) using similar methods, although different class averages were produced and five of them were used to form the initial model. This was used as an independent model from which to reconstruct an independent data set in order to determine the resolution of the reconstruction. (See Chapter 5, section 5.2.5: Resolution determination.)

4.2.3. Generation of symmetry file

In order to use these angles to produce a model in SPIDER (Frank *et al.*, 1996), it is necessary to create a symmetry file containing the correct angles for the icosahedral symmetry with MRC orientation, *i.e.*, three perpendicular two-fold axes concurrent with the *x*-, *y*- and *z*-axes. This was done by rotating the angles of the icosahedral symmetry file provided with SPIDER (Frank *et al.*, 1996), using MATLAB (MathWorks) to perform the calculations. The symmetry rotation matrix for the MRC object, from which the symmetry angles were calculated, was determined by applying a rotation matrix, with SPIDER Euler angles (φ = -90°, θ = 90°, ψ = -31.72°), to convert an icosahedron with MRC orientation into SPIDER orientation, then applying the rotation matrix with the Euler angles specified by the symmetry file, before applying the inverse of the first rotation matrix in order to rotate the object back to the MRC orientation. The overall angles of rotation for each equivalent orientation of the icosahedron, were extracted from this final rotation matrix and stored as a new symmetry file (Appendix B) specific to the MRC icosahedral orientation.

4.2.4. Initial model – EMAN

Another initial model was made by running the EMAN command, *starticos* (described in Chapter 4.1; Ludtke *et al.*, 1999), using 100 images per class in order to create an icosahedral starting model. This model was rotated by the EMAN Euler angles (58.27°, 270°, 0°) in order to orientate it in the MRC convention, then subjected to ten rounds of projection matching of the combined cryo-data set of HcRNAV109 images to improve the initially noisy starting model before refinement in separate defocus groups.

4.3. Results

The class averages produced by *K*-means classification are shown in Figure 4.1.

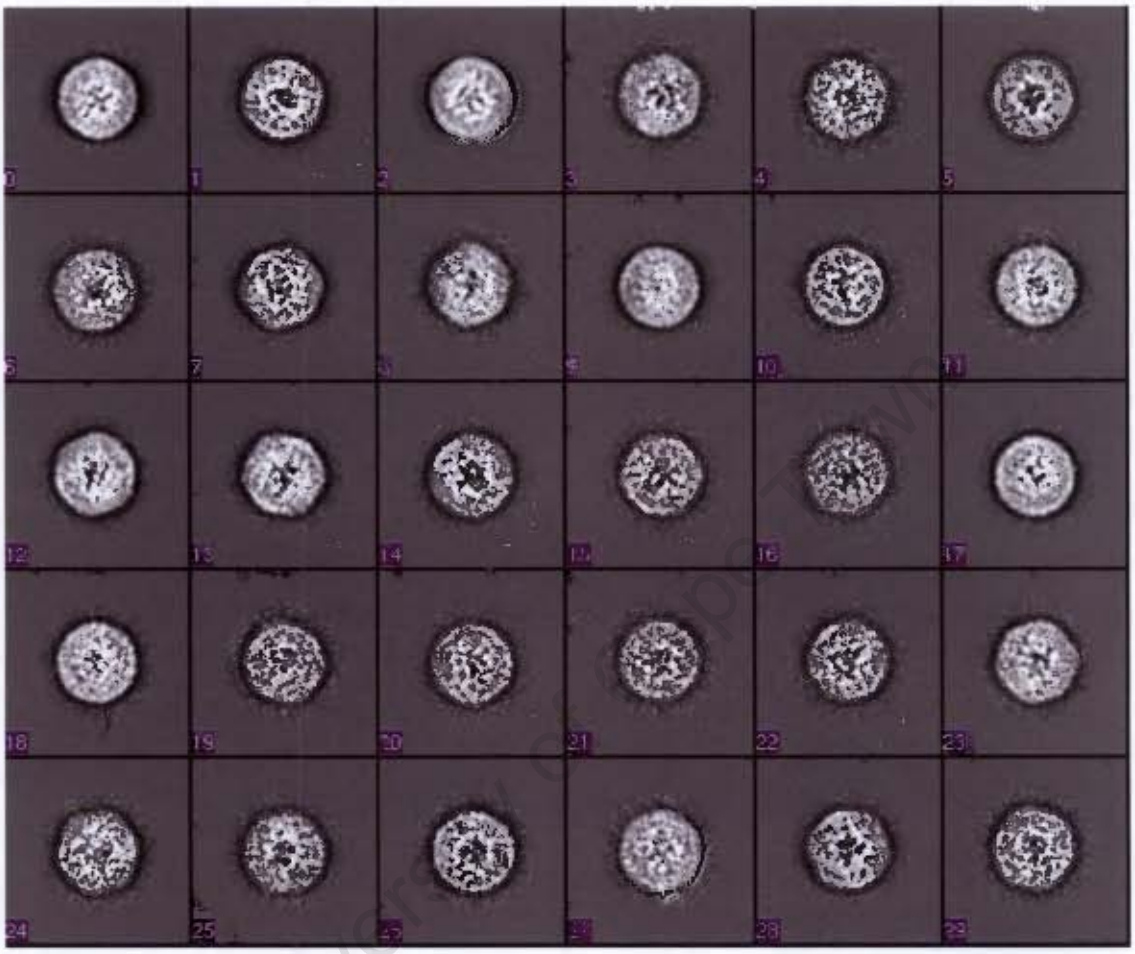

Figure 4.1: *Thirty class averages generated from one cycle of K-means classification with SPIDER (Frank et al., 1996) and viewed with EMAN (Ludtke et al., 1999).*

The orientation parameters determined by *refine* (Crowther *et al.*, 1996) are given with their corresponding weighted Crowther residuals in Table 4.1. Due to the poor contrast of the unstained, vitrified sample and therefore high SNR in the data (Fuller *et al.*, 1995), the results are not reliable, particularly near the symmetry axes, with their inaccurate orientations suggested by the sigma and weighted Crowther residuals. Therefore class averages had to be carefully picked. The starting models produced when multiple class averages and their respective orientations were used, were featureless due to inaccurate orientations causing averaging of symmetry-unrelated features.

Table 4.1: *Orientations of class averages and their weighted Crowther residuals as determined by refine (MRC; Crowther et al., 1996). The angles are given as SPIDER Euler angles.*

Class	ψ (°)	θ (°)	φ (°)	Weighted Crowther Residual
1	29	87	2	13.5
2	53	81	1	13.1
3	46	73	3	16.5
4	23	90	25	17.9
5	64	87	15	17.8
6	8	80	2	18.6
7	16	89	-6	17.8
8	37	81	13	18.9
9	8	87	-5	16.1
10	103	86	-9	17.4
11	126	88	-23	21.3
12	60	86	1	14.7
13	78	81	14	20.6
14	33	87	7	17.1
15	52	89	-23	15.5
16	104	90	-18	17.6
17	151	86	-10	18.9
18	168	89	-9	16.9
19	21	89	6	15.5
20	97	77	2	16.4
21	112	83	-14	18.8
22	25	87	-18	15.1
23	93	80	0	18.5
24	122	88	6	16.1
25	169	88	17	18.4
26	172	72	3	18.9
27	23	79	4	17.6
28	121	76	-1	20.8

Class average 19 is highlighted in Table 4.1 as it was used to produce the starting model. Although the Crowther residual is not the lowest for class 19, it showed consistency in the angle determined by the three statistics used: the Crowther residual, sigma and weighted Crowther residual. This class average, numbered 18 in Figure 4.1, shows clearly defined bumps, as did the model once it was constructed. The quality of this model was greatly improved by one round of refinement as shown in Figure 4.2a and b. This model resembles the Roseman starting model which is shown in Figure 4.2c.

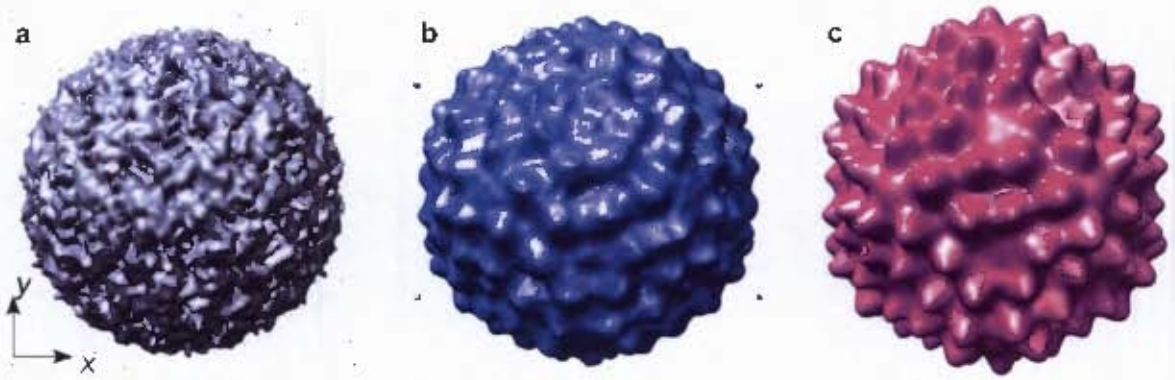

Figure 4.2: ***a)*** *Initial model generated by back-projection (SPIDER; Frank et al., 1996), with icosahedral symmetry, of class average 19 at the orientation angles determined by common lines (Crowther et al., 1996).* ***b)*** *the refined starting model after one cycle of projection matching with all the picked particles using SPIDER (Frank et al., 1996);* ***c)*** *The starting model produced by Alan Roseman (MRC LMB, Cambridge, UK) using common lines. Images produced with CHIMERA (Pettersen et al., 2004).*

The starting model generated by EMAN and the model after five rounds of refinement with the entire dataset are shown below (Figure 4.3). Figure 4.3c shows the differences in the two starting models generated by common lines and EMAN.

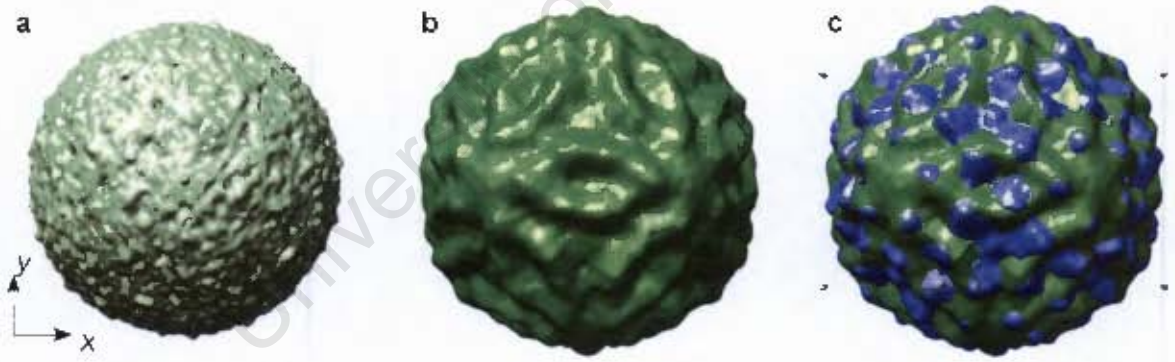

Figure 4.3: *The starting model generated by EMAN (Ludtke et al., 1999) (green)* ***a)*** *before refinement* ***b)*** *after one round of projection matching and* ***c)*** *compared to the model generated by common lines (blue). Visualised with CHIMERA (Pettersen et al., 2004).*

4.4. Discussion

The initial classes generated by *K*-means classification are not very distinct due to poor alignment of the images. This explains the poor success of the common lines method in determining the orientations of the particles. These errors were enhanced when more than one image was used to produce a model, forming featureless models. While further multi-reference alignment would have improved the class averages, and hence the orientations assigned to them, this proved unnecessary as a reasonable starting model was produced. The formation of a starting model from a single class average is possible due to the high symmetry of icosahedral viruses.

Both models produced by common lines are very similar due to the similar methods used in the construction of both of them. Although the Roseman model has sharper peaks, they are in similar positions to the common lines model and refined to the same final reconstruction.

While the initial model created by EMAN is featureless, the improvement of the model by iteration attests to the relative insignificance of the starting model, which is improved by refinement. The model at this stage is still distinctly different to the models formed by common lines (Figure 4.3c) and therefore is a good starting model for further resolution determination.

Chapter 5. Refinement

5.1. Introduction

Once an initial model is obtained, it can be used to refine the orientations assigned to the particles. This involves refining five parameters: three angles of rotation (ψ, θ and φ over the asymmetric unit; $\varphi \in [0°, 360°]$, $\theta \in [69.09°, 90°]$, $\psi \in [-31.17°, 31.17°]$) and two translation parameters (x and y). A variety of methods are used to refine these parameters for each image. These can be generally classified as following one of two model-based methods: common lines refinement, implemented in Fourier space (Zhou, 2003) or real space (van Heel *et al.*, 2000), and 3D projection matching, using cross-correlation or the polar Fourier transform (PFT; Baker & Cheng, 1996) to determine to which projection the images best align.

Both methods involve determination of the orientations of the particles based on relation to views of the model. In the *cross-common lines method,* the orientation of the particle relative to a small set of views of a model, with known orientations (usually projections), is determined by comparing the positions of pairs of common lines using the phase residual as a measurement as described in Chapter 4 (Crowther, 1971).

The *3D projection matching* algorithm, successfully used for many asymmetric particles as well as those with higher symmetry, involves the comparison of the compete set of projections of the asymmetric unit with each image to determine the correct orientations (ψ, θ) by assigning the orientation of the projection that matches the image best. This can be determined by cross-correlation of the image and each projection over a range of azimuthal angles (φ) and shifts (x, y) in order to align it to the reference projection (Frank, 2006). Alternatively this may be carried out for the Fourier transforms of the projections and images (Joyeux & Penczek, 2002). The PFT method increases the speed of this process by breaking this down into two steps. First, the rotationally invariant polar Fourier transforms of the projections and images are compared by cross-correlation to determine the best values of θ and ψ. Second, real space cross-correlation of the image with the appropriate projection is then used to refine the phase origin (x, y) and azimuthal orientation (φ) of the image (Baker & Cheng, 1996).

Crowther (1970b) determines the minimum number of evenly spaced views, n, needed to reconstruct an asymmetric particle with diameter D to a resolution of d:

(1) $n = \frac{\pi D}{d}$.

This is reduced 60-fold in the case of spherical viruses due to the icosahedral symmetry (Zhou & Chiu, 2003), allowing for reconstruction from very few views. However, the low SNR of the low dose images of the particles necessitates oversampling (Zhou & Chiu, 2003). Whereas only a few views are needed to reconstruct an icosahedral particle, many more projections are needed to accurately determine the orientations of the particles by projection matching. Small inaccuracies in the determination of particle orientations lead to loss of resolution of the model reconstructed from these views, particularly when symmetry is applied.

Once the orientations are determined, the particles (or the class averages of particles with the same orientations) are back-projected to reconstruct the 3D model. The three types of techniques used are weighted back-projection, Fourier reconstruction techniques and iterative algebraic reconstruction (Frank, 2006). With the *weighted back-projection* method, the density of the aligned views is back-projected from 2D into a 3D view. The summation of all the 3D volumes generated from the different views produces the 3D reconstruction. Weighting is necessary for the back-projection method because the conversion between polar and Cartesian coordinates in the algorithm leads to imbalance in the representation of spatial frequencies. While the weighting function is simple for evenly spaced projections, this becomes much more complex when the angular increments are not even (Frank, 2006). *Fourier reconstruction methods* use the projection theorem that the Fourier transforms of projections are central slices of the Fourier transform of the 3D model. Complications arise in interpolation as the transforms due to sampling the Fourier space at discreet points which may not coincide with the positions of the transforms of the images (Frank, 2006), but a fast and accurate Fourier interpolation method has been developed (Penczek *et al.*, 2004). *Iterative algebraic reconstruction* methods involve iteratively comparing the 2D images to projections of an estimated 3D volume and adjusting the volume to decrease the differences between them. These methods are quite flexible, producing high quality results, but require a large amount of computing power (Frank, 2006).

The various strategies used for the CTF correction differ on when in the reconstruction the CTF is corrected: directly to the micrograph, although separately in different areas, before processing (van Heel *et al.*, 2000); on particle averages collected in defocus pairs of the same image (Conway *et al.*, 1997); at the image level after the particles have been aligned and centred to models that have been

multiplied by a CTF (Zhou *et al.*, 1994); or at the 3D map level (Böttcher & Crowther, 1996; Penczek *et al.*, 1997). CTF correction of the whole micrograph or of the individual particles allows for refinement of all the data together, which increases the SNR by averaging many particles. The disadvantage is that the contrast, which is generated by the CTF and aids in alignment, is decreased. One of the problems with performing the CTF correction on the models, is that the increased SNR (which is due to the averaging of many images) is decreased when the reconstructions are carried out in separate defocus groups. This can be overcome by aligning the images to the projections of the CTF-corrected model (produced by combining the models of the different defocus groups with CTF correction; Penczek *et al.*, 1997), that has had the appropriate CTF applied.

There are a number of ways to correct for the CTF. The simplest correction is to simply reverse the sign of the negative phases so that they are all positive, leaving the amplitudes modified by the CTF. A better correction involves essentially dividing the transform of the image by the CTF and envelope function using a Wiener-like filter function so that the intensity of the transform is approximately constant except for the data points near the zeros in the CTF, which are excluded to avoid amplifying noise.

CTF correction is used to recover the amplitudes of the the Fourier transform of the images that are affected by the CTF. Since the Fourier transform of the image, $I(s)$, is the product of the Fourier transform of the object, $F^2(s)$ and the CTF,

$$(2) \quad I(s)=F^2(s)CTF^2(s)E^2(s)+N^2(s)$$

it would appear that, in the absence of noise ($N^2(s)$), the object should be recoverable by dividing the Fourier transform of the image by the CTF. However, noise is present and would be enhanced in regions where the CTF is small, at both the low- and high-spatial frequencies. More problematically, the CTF crosses zero producing areas where division is impossible. The solution, which takes this into account, is the Weiner filter (equation 3) which is plotted in Figure 5.1 (Penczek *et al.*, 1997):

$$(3) \quad W(s)=\frac{CTF^*(s)}{|CTF(s)|^2+1/SNR}$$

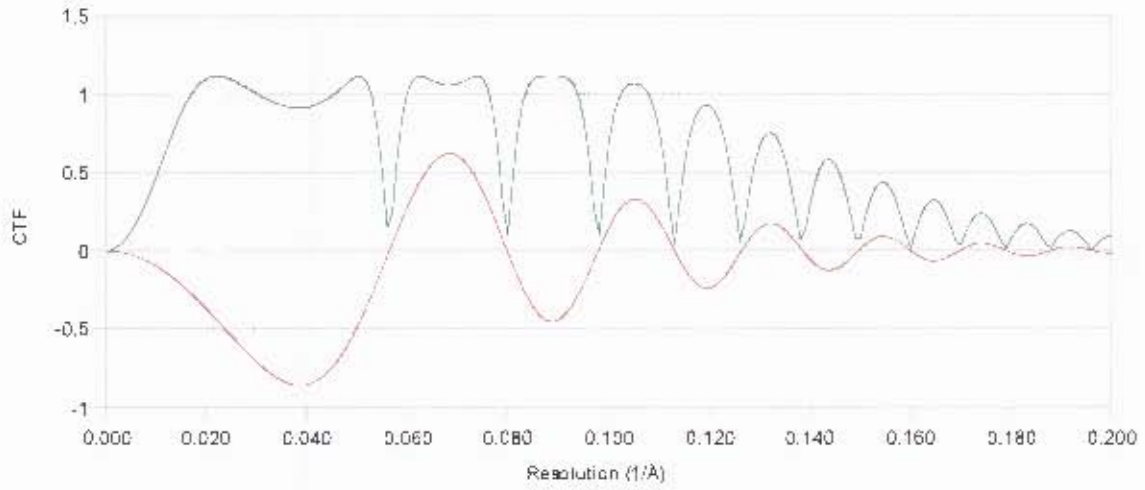

Figure 5.1: *Plot of the real parts of the calculated CTF (——) and Weiner (——) filter at 1.6 µm defocus with SNR = 5. The CTF was simulated by the web-based CTF simulation program (Jiang & Chiu, 2002). The Weiner function was calculated using equation 3.*

This produces images where the phases are all correct and the amplitudes are approximately constant at all spatial frequencies, except for the zeros. The loss of data at the frequencies of the zeros of the CTF is compensated for by combining data from different defocus groups that together cover the full spatial frequency range.

The resolution of a reconstructed 3D model is determined by comparing two independent reconstructions. Ideally the data should be split randomly into two independent data sets and the model reconstructed from two independent starting models to avoid model bias. The 2D class averages or 3D reconstructions are then compared in rings or shells in Fourier space. One measurement is the differential phase residual (DPR), which measures the amplitude-weighted mean phase difference between the reconstructions with the resolution cutoff determined to be where the DPR drops below 45° (Radermacher *et al.*, 1987). The more popular measurement is the Fourier ring or shell correlation (FRC or FSC), which measures the correlation of the reconstructions as a function of the spatial frequency (*i.e.* radius in Fourier space). Unfortunately there are two different criteria regularly used to determine the resolution cutoff. For one method the resolution is set at the spatial frequency at which the FSC drops below the threshold value, 0.5; the other method compares the curve to the 3-σ curve which indicates the expected correlation for pure

noise (Frank, 2002). The 3-σ criterion give resolution values that are consistently much higher than those determined using the 0.5 cutoff (Frank, 2006).

Since viruses are composed of both protein and nucleic acid, it is necessary to separate the protein and nucleic acid to determine the correct threshold for the volume using the densities of the two components as they have different densities and electron densities. Although the elastic scattering of electrons from nucleic acids is predicted to be more than twice as strong as that of proteins (Smith & Langmore, 1992), the symmetrisation of the particle decreases the electron density as the icosahedral symmetry is not necessarily extended to the nucleic acid.

While the predicted volumes of the protein and nucleic acid components may be calculated from their masses and predicted densities and the models thresholded at the appropriate level, this is a very inexact method. Small changes in the scaling and CTF can greatly affect the volume, which is very sensitive to the contour level (Baker *et al.*, 1999). Even with highly refined structures the calculated volume rarely matches the expected value. This may be due to limited resolution which produces a volume larger than expected (Conway *et al.*, 1996), or because the contributions of carbohydrate and lipid bound to the protein are excluded in the calculations because they are difficult to model. Additionally disordered portions of the structure may not be visible at high resolution. Because of this, models are often contoured at about 120% of the expected value (Baker *et al.*, 1999). An alternative method, which is more subjective and less easy to quantify, is to choose a volume at which there are no large disconnected areas and unexpected holes in the structure (Conway *et al.*, 1996). Of course, this is a guideline that can be disregarded when there are explanations for disconnected densities, such as disordered elements that may not be visible at the correct threshold (Conway *et al.*, 1996). A third, more objective method consists of setting the threshold level at 1.5 or 2 standard deviations above the background (Baker *et al.*, 1999).

5.2. Methods

5.2.1. Projection

In order for the images from the different defocus groups to align correctly, calculated CTFs of each defocus group were applied to the Fourier transform of the reference model, which was then Fourier transformed back into real space to produce models with the appropriate defocii.

Each model, was then projected at evenly spaced angles over the asymmetric unit ($\theta \in [69.09°, 90°]$; $\varphi \in [-31.17°, 31.17°]$). According to equation 1, for this 300 Å HcRNAV reconstruction, to achieve a 15 Å resolution, ~0.5 ($\frac{\pi.300}{15.60}$) projections would be required. To accurately determine the orientations of the particles, and to compensate for the low SNR, 29 projections with angular spacing of 5° were initially used for a redundancy of 55. After 25 iterations this was increased to 164 projections with an angular spacing of 2°, and then to an angular spacing of 1° producing 678 projections after 34 iterations so that the orientations of the particles may be more accurately determined.

5.2.2. Alignment

Multi-reference alignment, with AP SH from SPIDER (Frank *et al.*, 1996), was used to align the particles to the reference projections with cross-correlation coefficients determining the best reference and alignment. Rings of between 5 and 52 pixel radii (which was decreased to 47 pixels after 22 iterations) were compared, with a translation search range of 5 pixels in 1 pixel steps and no angular restrictions. Initially 90% of the images (rated according to their cross-correlation scores) in each defocus group were used to produce the class averages, and this was reduced to 80% after five iterations with a further restriction of only including images with a correlation coefficient greater than 1 700, which was increased to 2 000 after 36 iterations. The number of images were further decreased to 70% after 38 iterations and the threshold increased to 2 100. Class averages of the images were then formed for each reference projection.

The movement of individual images between classes was tracked in order to determine when the model had converged.

5.2.3. Back-projection

The class averages, which did not necessarily cover all the angles of the asymmetric unit that were originally projected, were then back-projected with BP 3F, applying icosahedral symmetry, to form a new model. When the number of projections was increased to 164, individual classes consisted of two to three particles at most, with most classes consisting of a single image. Therefore at this stage, the images were individually back-projected at the angles determined by the alignment.

5.2.4. CTF correction

Ideal CTFs were calculated for each defocus group using the defocus values calculated previously by CTFFIND3 (Mindell & Grigorieff, 2003) and Gaussian envelopes determined manually using CTFMATCH (Baxter, 2004). The SPIDER command, TF CTS (Frank & Penczek, 1995), corrects the model from each defocus group by the CTF for that defocus, weighted by the number of particles in the group to account for "noisy" models produced by defocus groups with fewer images. A SNR of 3 was used. All of the models were combined to give an overall CTF-corrected model, which was used to form the next set of projections for further refinement.

5.2.5. Resolution determination

The reconstruction was performed by splitting the images into two equally sized groups using two independent starting models to avoid model bias. The first model was created using common lines to determine the orientations of the particles with the MRC program, *refine* (Crowther *et al.*, 1996), and the second model was created using EMAN (Ludtke *et al.*, 1999) as described in Chapter 4. The FSC and DPR of the two independent reconstructions was determined in Fourier shells of 1 pixel^{-1}, using RF 3 from SPIDER (Frank *et al.*, 1996). A cutoff of 0.5 for the FSC and 45° for the DPR were used to determine the resolution. The resolution could also be improved by masking the RNA, which does not contain the same icosahedral symmetry as the capsid. This masked model was used for the refinement of models after 10 iterations.

5.2.6. Threshold

The model was thresholded to 0.0143, a value two standard deviations above the mean which produces a sensible looking model. The volume of the protein at this threshold was calculated with the RNA removed, by a mask with radius of 108 Å using the predicted protein density of 1.41 g. cm^{-1} (Fischer *et al.*, 2004).

5.2.7. Negative stain

Similar methods were used to obtain a 3D reconstruction from the negative stain data set. As the resolution was not expected to be better than 20 Å, CTF correction is unnecessary and therefore all of the particles are refined in one group (or split into two for resolution determination). The models

produced by the cryo-reconstruction were used as starting models. Due to the larger sampling, the models produced appeared smaller, and so a smaller radius (50 pixels) is used for the alignment and back-projection steps. All the images were used for the reconstruction in the first 5 refinements. The number of images was decreased to 90% for the next 10 iterations, to 80% for iterations 16 to 20 and to 70% for iterations 21 to 30. For iterations 31 to 35 the number of projections to which the images were aligned were increased from 29 to 164. The reconstruction was again carried out in two independent groups in order to determine the resolution by measuring the FSC and DPR.

5.3. Results

5.3.1. Cryo-EM

The projections of the model at 29 evenly spaced angles and some of the 164 angles is shown in Figure 5.2. It was to these projections that the images were aligned.

Figure 5.3 shows the images of one defocus group aligned to one of the projections and the class average from these projections that was used to for the back-projection step when angular steps of 5° were used.

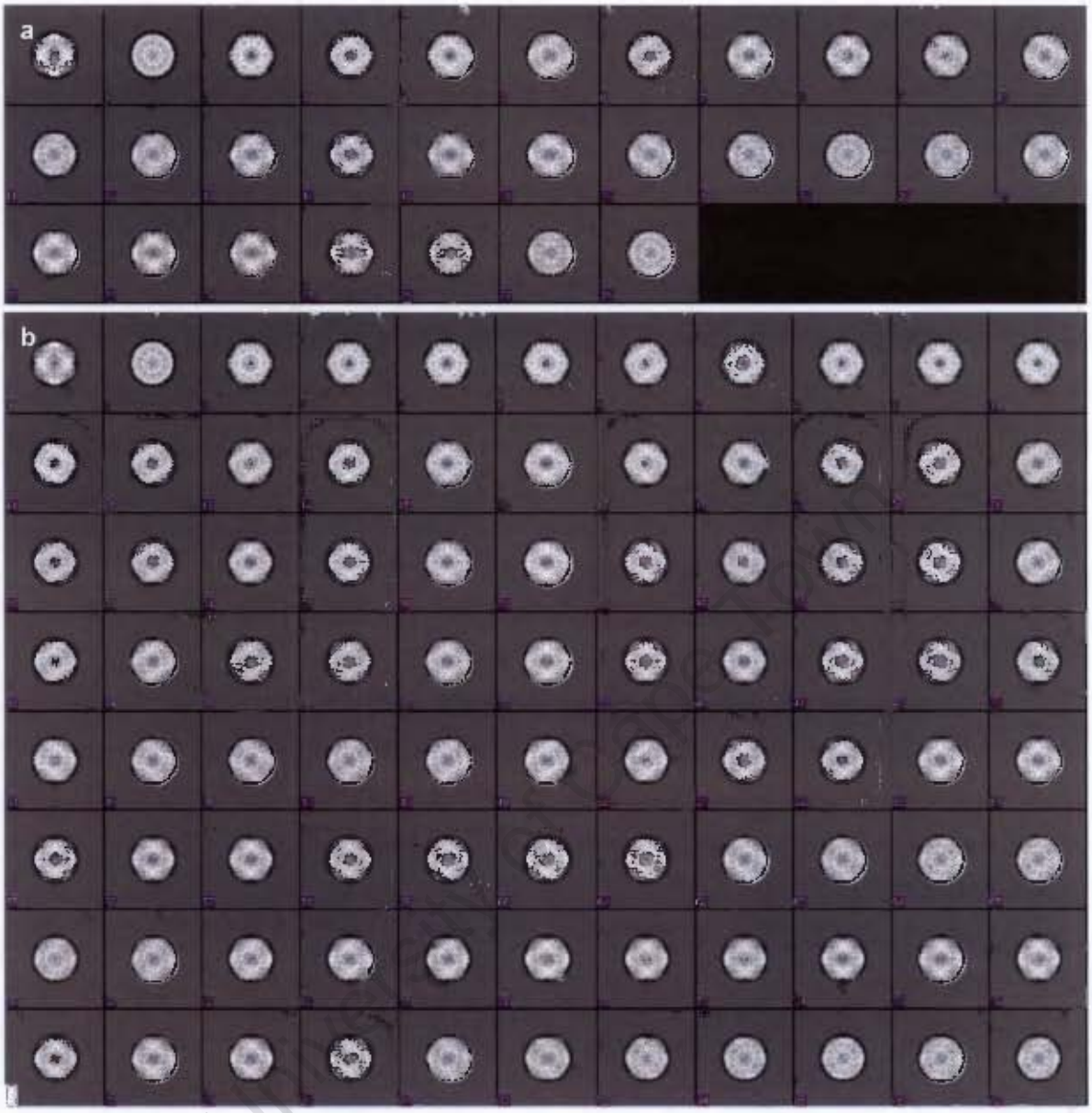

Figure 5.2: *Projections of the model over the asymmetric unit at,* ***a)*** *5° increments and* ***b)*** *2° increments. Projections were made with SPIDER (Frank et al., 1996) and visualised with EMAN (Ludtke et al., 1999).*

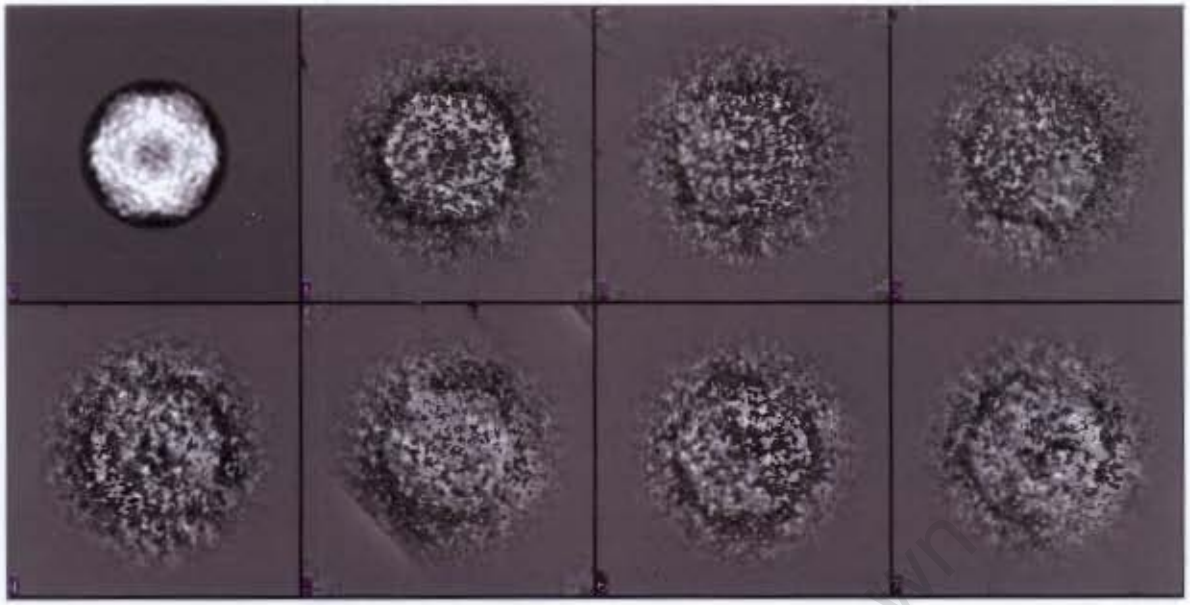

Figure 5.3: *One of the projections of the cryo reconstruction of HcRNAV109 (0), the corresponding class average (1) and particles (2-7) for the group with a defocus of 3.05 μm.*

The convergence of the refinement is shown in Figure 5.4 where the movement of images between classes is plotted for each iteration. The frequent increase in class changes is due to changes in the refinement parameters. The two largest peaks at 25 and 34 iterations are due to the increase in the number of projections used to align the images (from 29 to 164 and 164 to 678) as the number of classes is increased. After 10 iterations the RNA was masked in the reconstruction, after 17 iterations the CTFs were weighted by the number of images contained in them, at 20 iterations a threshold of 2 500 was applied to eliminate particles with low cross-correlation scores to the projections and at 37 iterations, the threshold, which had been reduced to 1700 with the increase in the number of projections at 34 iterations was increased to 2 000, at 41 iterations the threshold was again increased to 2100 and after 45 iterations the maximum number of images used in each defocus group was decreased to 70 to prevent domination by one defocus group. The reconstructions were not allowed to fully converge between parameter changes due to length of processing and time constraints.

This reconstruction ultimately led to the high resolution models, which are shown in Figure 5.5.

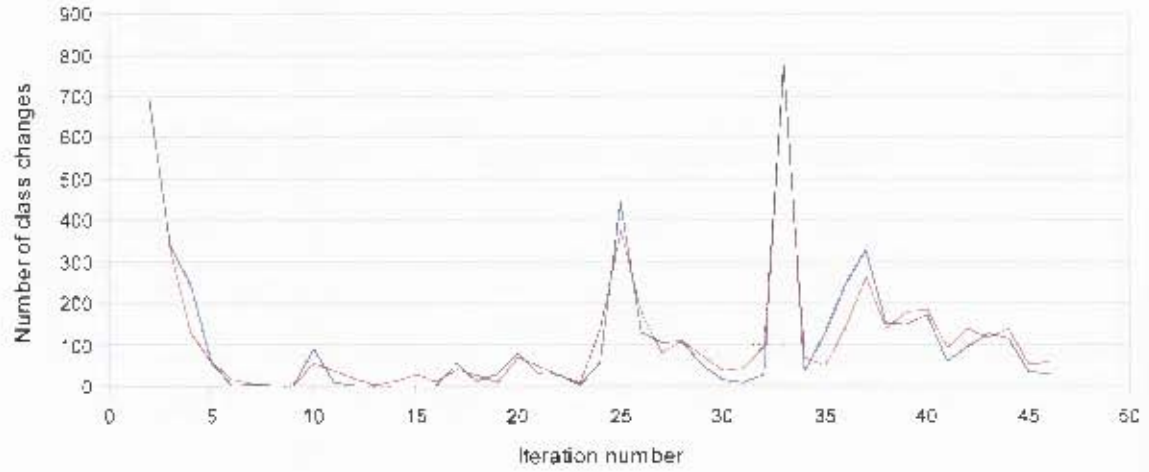

Figure 5.4: Convergence of the two independent cryo reconstructions, (——) and (——), of HcRNAV109 as indicated by the number of class changes in the two independent reconstructions.

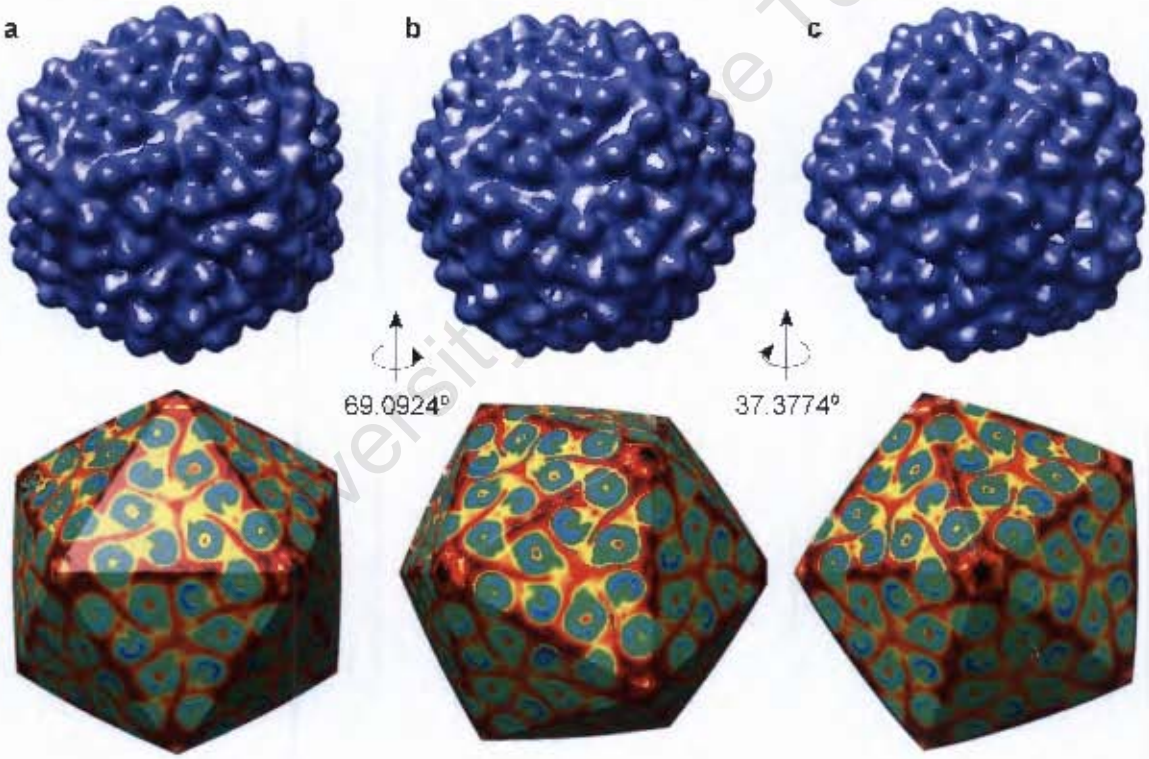

***Figure 5.5: a)** Two-fold, **b)** three-fold and **c)** five-fold surface views of the cryo reconstruction of HcRNAV with the electron density mapped onto an icosahedral surface below them. The electron density increases from black, red, green, aquamarine to blue. The reconstruction was performed with SPIDER (Frank et al., 1996) and the images created with CHIMERA (Pettersen et al., 2004).*

The two independent reconstructions were used to calculate the resolution using the FSC and DPR as measurements of the resolution. Significant improvements in resolution were seen when the images were aligned to the CTF corrected model rather than the individual defocus group models (from 27.3 Å to 19.8 Å) and when the angle step size for producing the projections was decreased from 5° to 2° (to 18.5 Å). Decreasing the angle step size to 1° improved the resolution slightly to 18.1 Å (Figure 5.6).

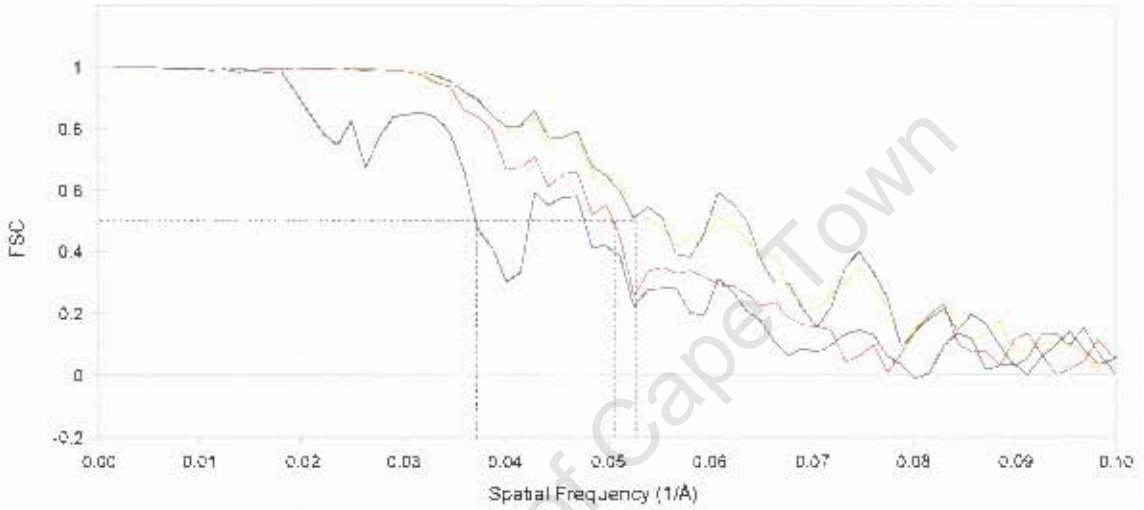

Figure 5.6: *FSC plots with the 0.5 resolution criterion marked for: (——) the models refined in separate defocus groups, (——) the images aligned to projections of the merged model with 5° angular increments, (——) images aligned to projection at 2° angular increments and (——) images aligned to projection at 1° angular increments.*

Figure 5.7 shows that there is a final resolution of 22.6 Å according to the DPR criterion of 45° and 18 Å according to the FSC criterion of 0.5.

The threshold value of double the standard deviation above the mean that was used in visualising the models. This produces models where the capsid portion (with the RNA masked out) has a volume of 1.037×10^7 Å^3 (219.4×10^3 voxels), which is larger by 22% to the predicted volume of 180 subunits of the 38 kDa protein of 8.055×10^6 Å^3 (using protein density = 1.41 g.cm^{-3}).

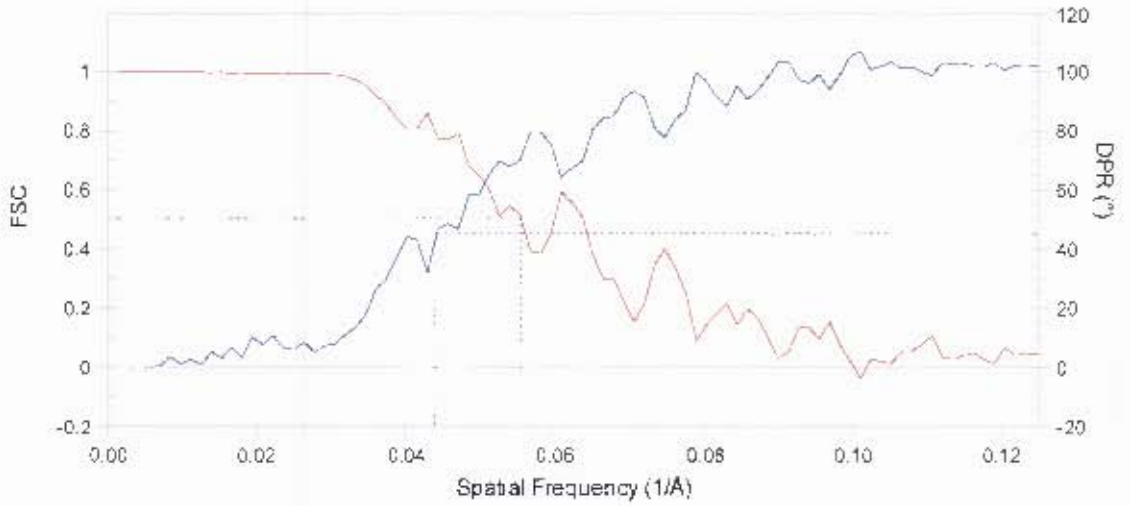

Figure 5.7: FSC (——) and DPR (——) of the final negative stain reconstruction indicating resolution values of 18.1 Å and 22.6 Å respectively.

5.3.2. Negative stain

Due to the starting models that were close to the true model, the reconstruction converged easily (within 30 iterations, Figure 5.8) to the model shown in Figure 5.9 with a radius of 188 Å. The increases in the convergence plot at 16 and 21 iterations are due to decreases in the number of images used in the reconstructions from 90% to 80% to 70%.

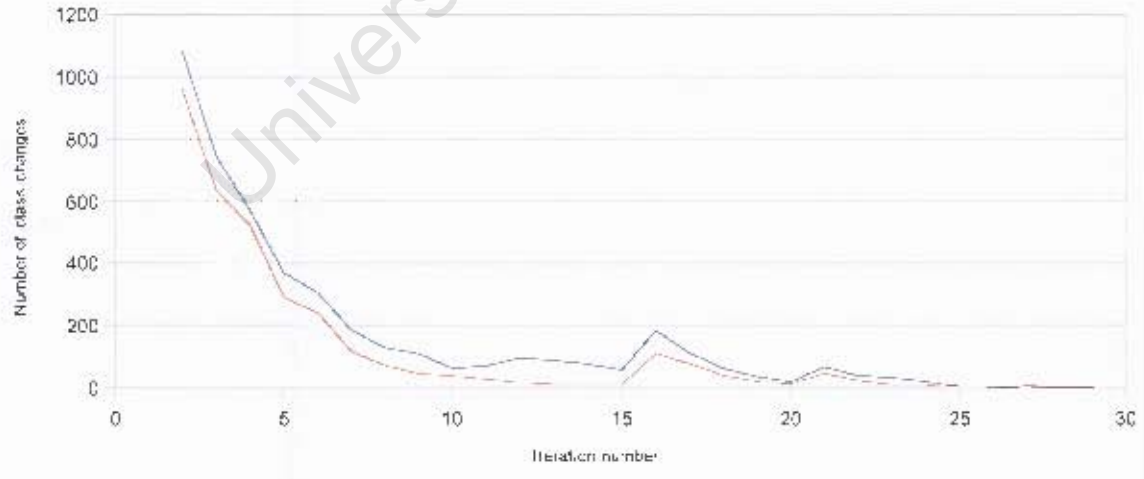

Figure 5.8: *Convergence of the two independent negative stain models, (——) and (——), of HcRNAV109 as indicated by the number of class changes in the two independent reconstructions*

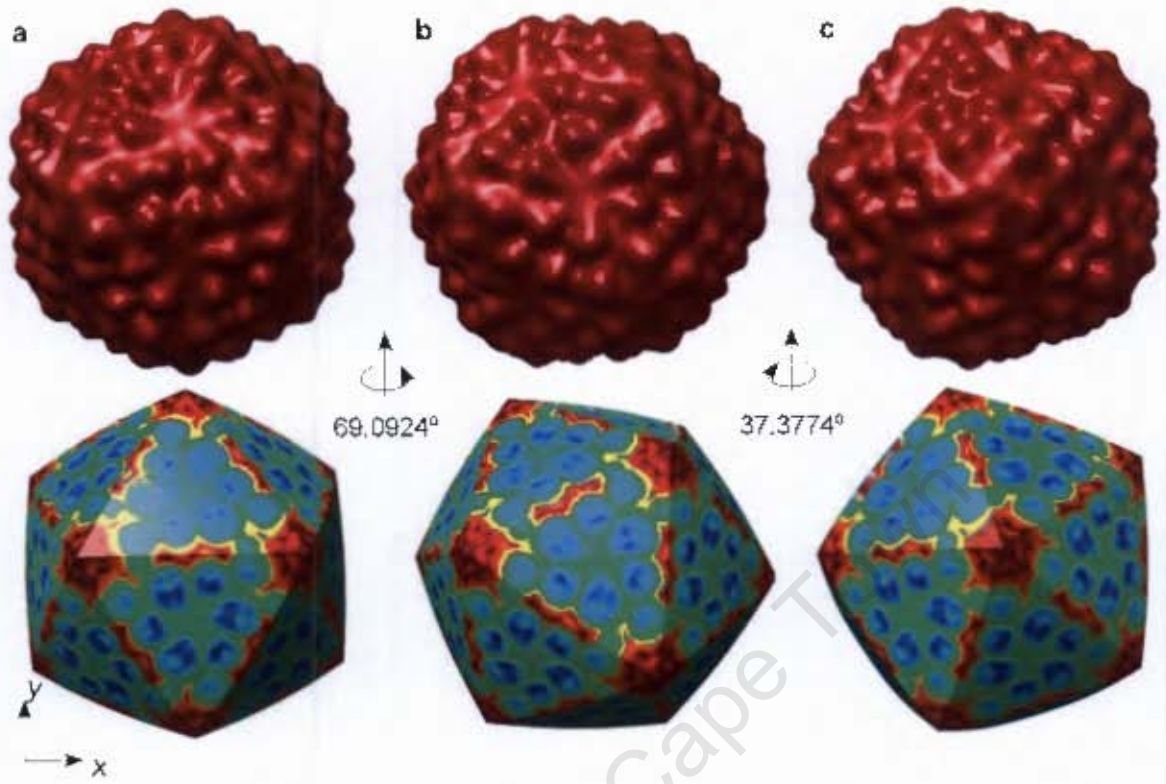

Figure 5.9: *Surface views and icosahedral slices of the electron density of the final negative stain model of HcRNAV109* ***a)*** *down the two-fold axis,* ***b)*** *down and three-fold axis and c) down the five-fold axis. Reconstructed with SPIDER (Frank et al., 1996) and visualised with CHIMERA (Pettersen et al., 2004).*

The FSC shown in Figure 5.10 indicates a final resolution of 25 Å that did not improve with a decrease in angular step size.

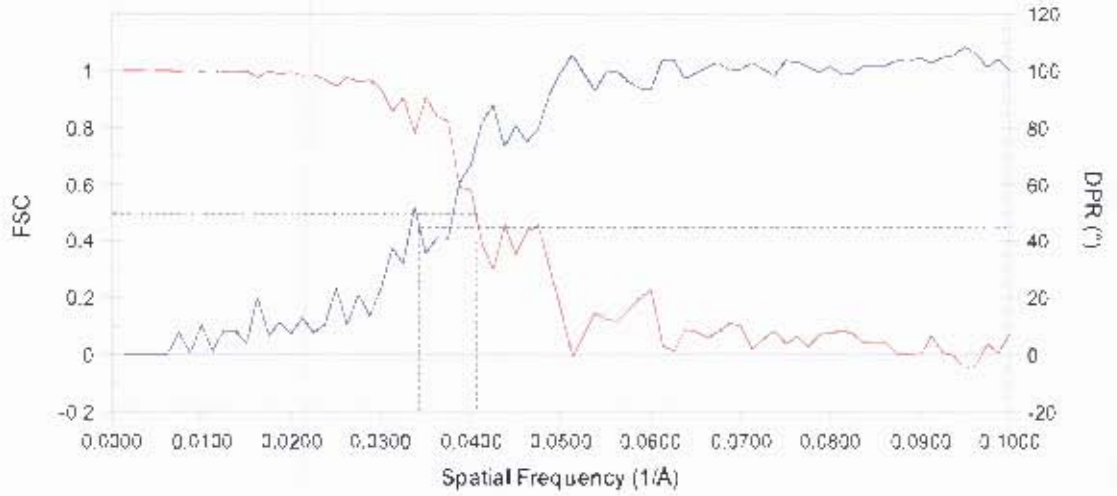

Figure 5.10: *FSC (——) and DPR (——) of the final negative stain reconstruction of HcRNAVI09 indicating resolution values of 25 Å and 26 Å respectively.*

5.4. Discussion

One of the problems with performing the CTF correction on the models, is that the increased SNR which is generated by combining many images, is decreased when the reconstructions are carried out in separate defocus groups. This is overcome by aligning the images to the projections of the CTF corrected model, which has had the appropriate CTF applied. A large increase in resolution (from 27.3 Å to 19.8 Å resolution) was achieved when this was first implemented and the reconstruction was therefore repeated with this feature from the start.

The number of images used for the reconstruction is decreased gradually in order to slowly remove the worst images based on their correlation to the projections in order to improve the model. While more images increase the SNR in the class averages that are used to produce the model, the deviation from the 3D structure by poorly aligned images and deformed or overlapping particles distorts the class averages used to form the models, thereby decreasing the resolution. This has been shown by Liu *et al.* (2007) where a resolution of 7.9 Å was achieved for a reconstruction from 284 particles chosen from a data set of 4 865 images originally collected (Liu *et al.*, 2007).

This strategy, together with the increase in the number of projections led to a final resolution of 18.1 Å. While this is markedly better than the 25 Å resolution obtained by reconstruction of the

negative stain data set, it is at a worse resolution than is often achieved for cryo-reconstruction of virus particles. Unfortunately, due to lack of material, this resolution could not be improved by collecting more data. This also prevented the use of X-ray crystallography to provide a detailed, high resolution structure of HcRNAV109.

Initially the two independent models appeared to be very different when aligned. Closer examination revealed that this was due to opposite handedness of the molecules (Figure 5.11). Because the images are projections of the molecule, it is impossible to tell apart the top and bottom of the particles from the images. Therefore two mirror-image reconstructions are possible. Tilting or metal-shadowing experiments are required to determine the handedness of the virus and thus it was

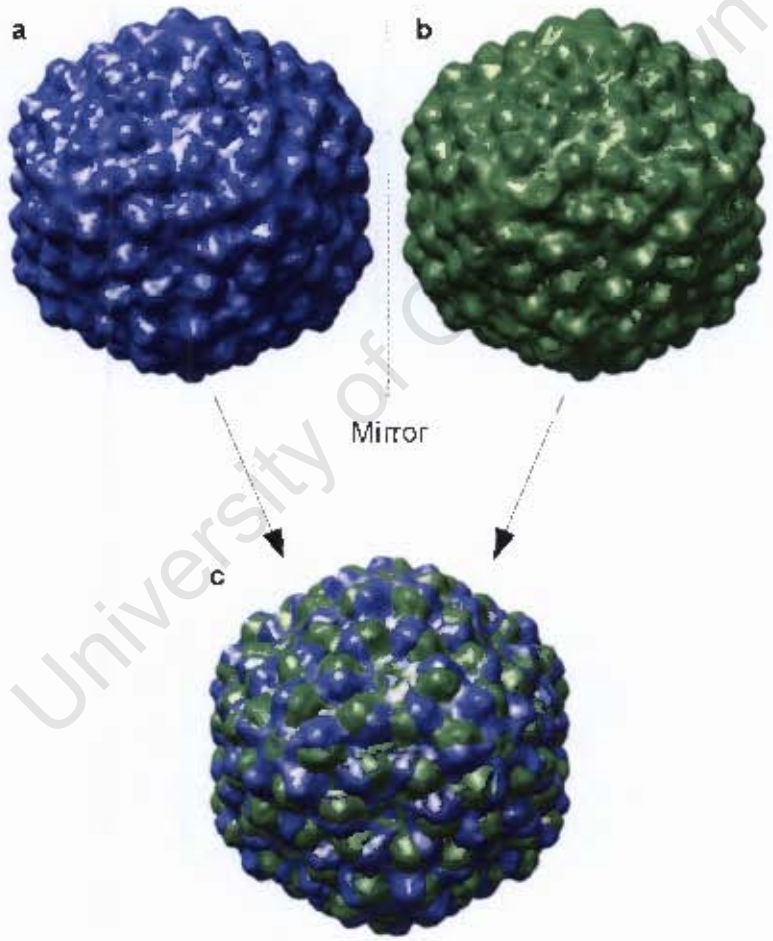

Figure 5.11: *Handedness of the two independent cryo reconstructions of HcRNAV109 after ten iterations.* ***a)*** *The reconstruction from the common lines starting model.* ***b)*** *The reconstructions from the EMAN starting model.* ***c)*** *Both reconstructions aligned to one another. Visualised with Chimera (Pettersen et al., 2004).*

necessary to arbitrarily chose one of the models. The common lines reconstruction was arbitrarily chosen and so the reconstruction from the EMAN starting model was mirrored to reflect identical handedness (Figure 5.11).

Examination of the arrangement of subunits of the model reveals that the virus is a T=3 virus, as can be seen when a cage representing T=3 symmetry is imposed (Figure 5.12a). The 180 quasi-equivalent subunits that are predicted by the symmetry and appear as 180 "bumps" on the surface of the capsid, is verified by the volume calculation. The different subunits and pentamers are clearly visible when the model is coloured radially (Figure 5.12b and c).

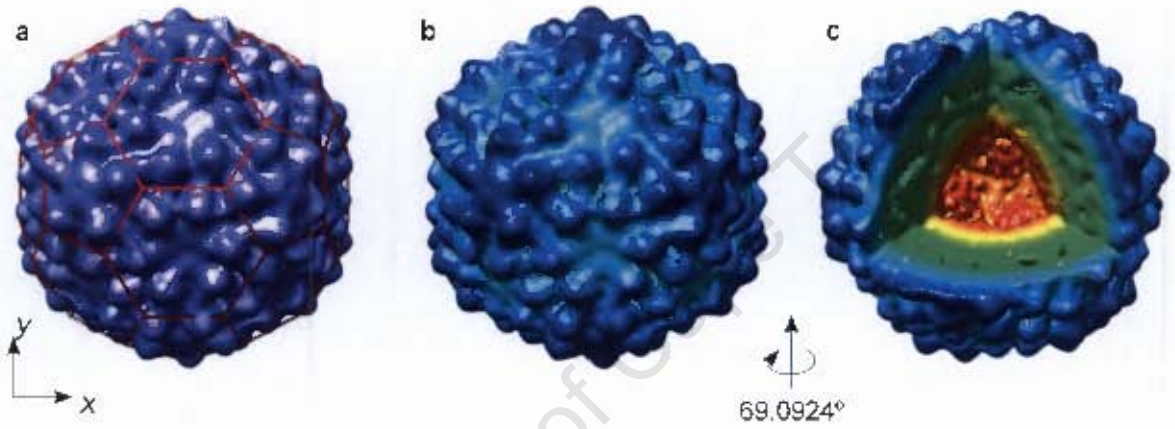

Figure 5.12: *Various views of the cryo reconstructed HcRNAV109: **a)** with a T=3 symmetry cage, **b)** coloured by radius to clarify symmetry elements and **c)** with an octant cut away to reveal centre. Images created with Chimera (Pettersen et al., 2004).*

A cross-section view of the model (Figure 5.13a) coloured according to the electron density (value of the voxels) reveals the RNA inside the capsid, distinct from the capsid proteins. Although RNA usually has a much higher electron density than protein, the icosahedral averaging of the RNA decreases the apparent electron density in the model. Areas of higher electron density between the RNA and the capsid indicate areas of RNA that are more icosahedrally ordered that are most likely where the RNA binds to the protein. By slicing the electron density higher along the z-axis Figure 5.13b, it is revealed that these ordered-RNA regions occur at the three-fold axes. The N-terminus of the capsid protein contains an arginine-rich region (12 of the 18 amino acids from residue 22 to 39) which is most likely the RNA binding region. This suggests that N-terminal is positioned towards the interior of the virus.

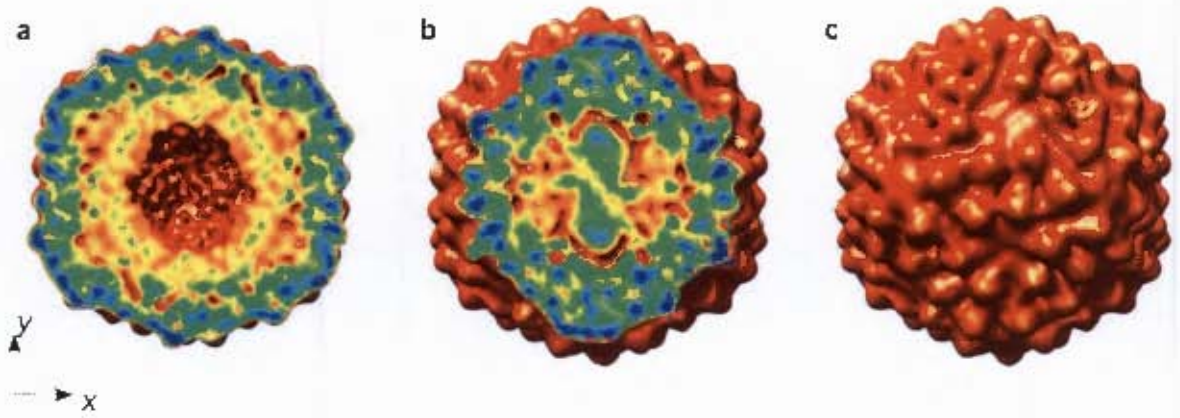

Figure 5.13: *View of the cryo model of HcRNAV109 down the two-fold axis coloured by electron density and sliced at* ***a)*** *half way,* ***b)*** *higher up and* ***c)*** *surface view in order to show the separate electron density of the protein and RNA. Viewed with Chimera (Pettersen et al., 2004).*

The large cavity in the centre of the RNA could be an artefact because the noise in the reconstruction of icosahedral particles is greatest here (Cheng *et al.*, 1994). The blobs of unconnected density are due to noise that has been symmetrised and filtered.

The trimers making up the capsid are visible in Figure 5.14 where there are distinct interactions between the the quasi-equivalent sub-units. The two slightly different *a* and *c* subunits alternate to form the hexamer around the three-fold axis.

The negative stain and cryo reconstructions appear outwardly similar, although the negative stain is at a lower resolution. However, slicing the two reconstructions reveals that the distinct separation of RNA from the capsid protein visible in the cryo-reconstruction is not visible in the negative stain reconstruction Figure 5.15. This is due to the stain not entering the viral shell to increase the electron density for it to be visible. The negative stain model shows greater electron density at the five-fold axes, indicated by blue in Figure 5.15a. This is less distinct in the cryo-reconstructions suggesting that it is an artefact due to the high symmetry at those positions.

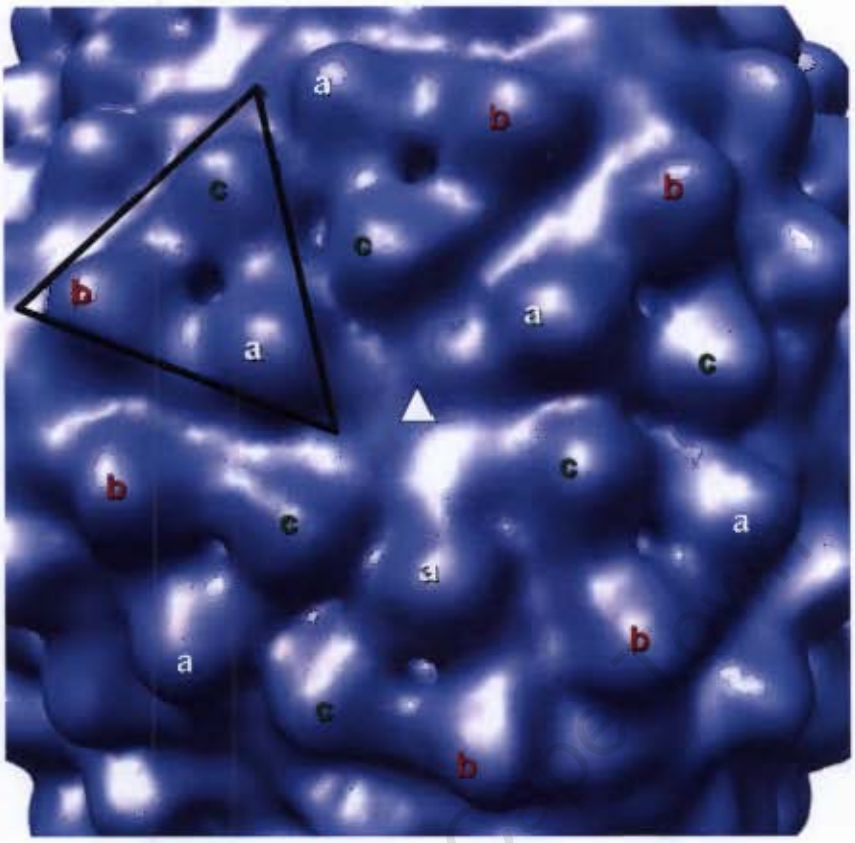

Figure 5.14: *Close up view of the capsid of the cryo reconstructed HcRNAV109 viewed down the three-fold axis displaying the capsid trimers (marked by the large black triangle. The different quasi-equivalent subunits are labelled* ***a), b)*** *and* ***c)****. Visualised with Chimera (Pettersen et al., 2004).*

This structure differs markedly from the structure predicted by Nagasaki *et al.* (2005) using homology modelling to the black beetle virus. Unlike the homology model which suggested one morphological subunit per capsid trimer, leading to the suggestion that this was a T=1 virus, this EM reconstruction clearly reveals that this is a T=3 virus. Each protein therefore produces an individual morphological subunit that adds up to the 180 expected for a T=3 virus (Figure 5.14).

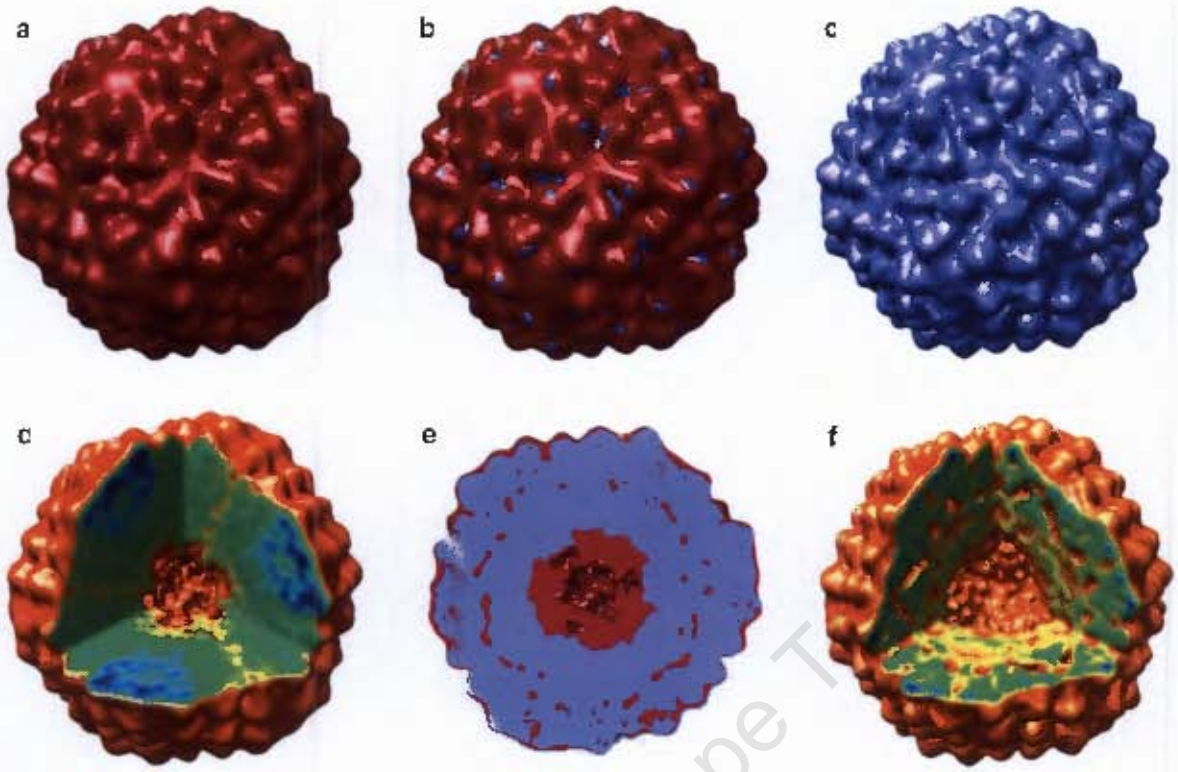

Figure 5.15: *Comparison of negative stain and cryo-reconstructions of HcRNAV109. Surface views of* ***a)*** *the negative stain reconstruction,* ***b)*** *both cryo- and negative stain reconstructions aligned and* ***c)*** *the cryo reconstruction are placed above their respective cut-away views of* ***d)*** *the negative stain and* ***f)*** *the cryo-reconstruction coloured by the electron density and* ***e)*** *the sliced view of both reconstructions aligned. Images made with CHIMERA (Pettersen et al., 2004)*

Black beetle capsid protein is a poor choice for a homology model as it only has 10% sequence homology to the HcRNAV capsid and no discernible predicted structure homology (Chapter 1, section 1.1 Heterocapsa circularisquama RNA virus), although the positions of the variable regions of the capsid were mostly located on the exterior of the virus which supported the hypothesis that variation in the capsid determines the strain specificity of the HcRNAV infection. Unfortunately, the resolution of this EM reconstruction does not allow for prediction of the tertiary structure of individual capsid proteins and further structural studies are needed in order to determine the positions of the variable regions of the capsid and the cause of the strain specificity of HcRNAV.

Chapter 6. Conclusions

Two 3D reconstructions of HcRNAV109 were produced for the two different methods of sample preparation: negative stain and cryo-EM. The final negative stain reconstruction was obtained using 2 457 of the 3 511 images originally picked for a final resolution of 25 Å. In total 1 275 of the original 2 593 particles, distributed amongst 19 defocus groups from the cryo-dataset were used to produce a model with a resolution of 18.5 Å. CTF correction was performed on the models from the separate defocus groups of the cryo-dataset to form one CTF-corrected reconstruction.

Two adjustments to the initial method were found to greatly improve the resolution of the reconstruction: aligning the particles to projections of the CTF-corrected model (with the appropriate defocus applied) rather than the individual models produced in the defocus groups; and increasing the number of projections to which the images were aligned by decreasing the angular step size. The increase in resolution due to more accurate determination of particle orientation by aligning them to more projections, is limited by the quality of the data. For the negative stain reconstruction, an increase in the number of projections did not increase the resolution of the reconstruction. While the resolution of the cryo-EM reconstruction improved when the angular step was decreased from 5° to 2°, further decrease of the angular step to 1° only improved the resolution slightly. The resolution of the cryo-reconstruction was limited by the number of images of HcRNAV109 particles that were collected due to limited sample with a low concentration of virus particles.

Reconstruction from the two independent starting models for the cryo-data illustrates a shortcoming of the method in that the two models have opposite handedness. This can only be resolved by obtaining additional information. Metal-shadowing or tilt experiments can be used to determine the handedness by EM. Alternatively if an atomic model becomes available where the handedness is restricted by the chirality of the natural L-amino acids or right-handed α-helices, it could be docked into the reconstructions to determine the correct structure. The correct model could, therefore, be the mirror-image of the one presented here.

This reconstruction reveals that HcRNAV109 forms a T=3 virus with each of the 180 subunits in the structure representing a separate morphological subunit. This number of subunits is consistent

with the volume calculated from the capsid component of the reconstruction. The interactions of the three quasi-equivalent subunits to form a trimer are revealed in the cryo reconstruction.

While the surface features are present in the negative stain reconstruction, no internal structure is visible due to the stain only coating the exterior of the virus. In contrast, in the cryo reconstruction of HcRNAV109, the capsid protein and RNA are clearly separated by a trough of low electron density. Although nucleic acid has a higher electron density than protein, the RNA of the virus reconstruction has a lower electron density due to averaging by the icosahedral symmetry that is imposed on the unsymmetrical RNA. Areas of more ordered RNA, where the RNA and protein interact more closely, are visible around the three-fold symmetry axes where the electron density is greater. This suggests that the RNA interacts with the capsid protein at the three-fold axes. Examination of the protein sequence reveals that the protein contains an arginine rich sequence near the N-terminal that most likely binds the RNA.

The homology modelling of HcRNAV34 and HcRNAV109 to BBV by Nagasaki *et al.* (2005) to predict the structure of the capsid proteins and determine the position of the variable regions of the capsid proteins has been revealed to be incorrect. Unlike BBV where the three quasi-equivalent proteins come together to form one morphological subunit, HcRNAV109 has three distinct "bumps" for the three subunits. Unfortunately, the resolution of the reconstruction does not allow for prediction of the tertiary structure of individual capsid proteins and thus nothing can be concluded about fold of the capsid protein and the positions of the variable regions of the sequence that are predicted to dictate the strain specificity of HcRNAV.

The lack of homology of the the capsid protein to any known virus does not discount the possibility that HcRNAV109 may also contain the β-jelly roll fold that is ubiquitous in virus capsid proteins, particularly ssRNA viruses, as the same topology is present in a wide range of capsid protein with no discernible sequence homology. However, no other related structures of marine algae viruses have been solved, so it is possible that capsid consists one of the other folds observed in virus, or an entirely novel fold. Higher resolution EM studies or X-ray crystallography structural studies are needed to determine the secondary and tertiary structure of HcRNAV.

References

Auer, M. (2000). Three-dimensional electron cryo-microscopy as a powerful structural tool in molecular medicine. *Journal of Molecular Medicine*, *78*, pp. 191-202.

Baker, T. S. & Cheng, R.H. (1996). A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *J. Struct. Biol*, *116*, pp. 120-130.

Baker, T. S., Drak, J. & Bina, M. (1988). Reconstruction of the three-dimensional structure of simian virus 40 and visualization of the chromatin core. *Proc Natl Acad Sci U S A*, *85*, pp. 422-426.

Baker, T. S., Drak, J. & Bina, M. (1989). The capsid of small papova viruses contains 72 pentameric capsomeres: direct evidence from cryo-electron-microscopy of simian virus 40. *Biophys J*, *55*, pp. 243-253.

Baker, T. S., Newcomb, W. W., Booy, F. P., Brown, J. C. & Steven, A.C. (1990). Three-dimensional structures of maturable and abortive capsids of equine herpesvirus 1 from cryoelectron microscopy. *J Virol*, *64*, pp. 563-573.

Baker, T. S., Newcomb, W. W., Olson, N. H., Cowsert, L. M., Olson, C. & Brown, J.C. (1991). Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. *Biophys J*, *60*, pp. 1445-1456.

Baker, T. S., Olson, N. H. & Fuller, S.D. (1999). Adding the Third Dimension to Virus Life Cycles: Three-Dimensional Reconstruction of Icosahedral Viruses from Cryo-Electron Micrographs. *Microbiology and Molecular Biology Reviews*, *63*, pp. 862-922.

Baldwin, P. R. & Penczek, P.A. (2006). The Transform Class in SPARX and EMAN2. *Journal of Structural Biology*, *157*, pp. 250-261.

Bamford, D. H., Burnett, R. M. & Stuart, D.I. (2002). Evolution of viral structure. *Theor Popul Biol*, *61*, pp. 461-470.

Bamford, D. H., Gilbert, R. J. C., Grimes, J. M. & Stuart, D.I. (2001). Macromolecular assemblies: greater than their parts. *Curr. Opin. Struct. Biol*, *11*, pp. 107-113.

Baxter, B. (2004). CTFMatch 1.0.

Bellare, J. R., Davis, H. T., Scriven, L. E. & Talmon, Y. (1988). Controlled environment vitrification system: An improved sample preparation technique. *Journal of Electron Microscopy Technique*, *10*, pp. 87-111.

Belnap, D. M., Olson, N. H., Cladel, N. M., Newcomb, W. W., Brown, J. C., Kreider, J. W., Christensen, N. D. & Baker, T.S. (1996). Conserved features in papillomavirus and polyomavirus capsids. *J. Mol. Biol*, *259*, pp. 249-263.

Benson, S. D., Bamford, J. K., Bamford, D. H. & Burnett, R.M. (1999). Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell*, *98*, pp. 825-833.

Bhella, D., Goodfellow, I. G., Roversi, P., Pettigrew, D., Chaudhry, Y., Evans, D. J. & Lea, S.M. (2004). The structure of echovirus type 12 bound to a two-domain fragment of its cellular attachment protein decay-accelerating factor (CD 55). *J Biol Chem*, *279*, pp. 8325-8332.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. *Nature*, *276*, pp. 362-368.

Boisset, N., Penczek, P., Taveau, J., Lamy, J., Frank, J. & Lamy, J. (1995). Three-dimensional reconstruction of Androctonus australis hemocyanin labeled with a monoclonal Fab fragment. *J Struct Biol*, *115*, pp. 16-29.

Booth, C. R., Jiang, W., Baker, M. L., Zhou, Z. H., Ludtke, S. J. & Chiu, W. (2004). A 9 angstroms single particle reconstruction from CCD captured images on a 200 kV electron cryomicroscope. *J Struct Biol*, *147*, pp. 116-127.

Böttcher, B. & Crowther, R. (1996). Difference imaging reveals ordered regions of RNA in turnip yellow mosaic virus. *Structure*, *4*, pp. 387-394.

Böttcher, B., Kiselev, N. A., Stel'Mashchuk, V. Y., Perevozchikova, N. A., Borisov, A. V. & Crowther, R.A. (1997a). Three-dimensional structure of infectious bursal disease virus determined by electron cryomicroscopy. *J Virol*, *71*, pp. 325-330.

Böttcher, B., Wynne, S. A. & Crowther, R.A. (1997b). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature*, *386*, pp. 88-91.

Bowman, V. D., Chase, E. S., Franz, A. W. E., Chipman, P. R., Zhang, X., Perry, K. L., Baker, T. S. & Smith, T.J. (2002). An antibody to the putative aphid recognition site on cucumber mosaic virus recognizes pentons but not hexons. *J Virol*, *76*, pp. 12250-12258.

Brenner, S. & Horne, R. (1959). A negative staining method for high resolution electron microscopy of viruses. *Biochim Biophys Acta*, *34*, pp. 103-110.

Bubeck, D., Filman, D. J., Cheng, N., Steven, A. C., Hogle, J. M. & Belnap, D.M. (2005). The structure of the poliovirus 135S cell entry intermediate at 10-angstrom resolution reveals the location of an externalized polypeptide that binds to membranes. *J Virol*, *79*, pp. 7745-7755.

Butcher, S. J., Bamford, D. H. & Fuller, S.D. (1995). DNA packaging orders the membrane of bacteriophage PRD1. *EMBO J*, *14*, pp. 6078-6086.

Canady, M. A., Tihova, M., Hanzlik, T. N., Johnson, J. E. & Yeager, M. (2000). Large conformational changes in the maturation of a simple RNA virus, nudaurelia capensis omega virus (NomegaV). *J Mol Biol*, *299*, pp. 573-584.

Caspar, D. L. D. (1956). Structure of a bushy stunt virus. *Nature*, *177*, pp. 475-476.

Caspar, D. L. & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol*, *27*, pp. 1-24.

Chapman, M. S. & Liljas, L. (2003). Structural folds of viral proteins. *Adv Protein Chem*, *64*, pp. 125-196.

Cheng, R. H., Reddy, V. S., Olson, N. H., Fisher, A. J., Baker, T. S. & Johnson, J.E. (1994). Functional implications of quasi-equivalence in a T= 3 icosahedral animal virus established by cryo-electron microscopy and X-ray crystallography. *Structure*, *2*, pp. 271-282.

Chipman, P. R., Agbandje-McKenna, M., Kajigaya, S., Brown, K. E., Young, N. S., Baker, T. S. & Rossmann, M.G. (1996). Cryo-electron microscopy studies of empty capsids of human parvovirus B19 complexed with its cellular receptor. *Proc Natl Acad Sci U S A*, *93*, pp. 7502-7506.

Choi, H. K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G. & Wengler, G. (1991). Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature*, *354*, pp. 37-43.

Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J. & Steven, A.C. (1997). Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature*, *386*, pp. 91-94.

Conway, J. F., Trus, B. L., Booy, F. P., Newcomb, W. W., Brown, J. C. & Steven, A.C. (1996). Visualization of three-dimensional density maps reconstructed from cryoelectron micrographs of viral capsids. *J Struct Biol*, *116*, pp. 200-208.

Crowther, R. A. (1971). Procedures for Three-Dimensional Reconstruction of Spherical Viruses by Fourier Synthesis from Electron Micrographs. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *261*, pp. 221-230.

Crowther, R. A. (2004). Viruses and the development of quantitative biological electron microscopy. *IUBMB Life*, *56*, pp. 239-248.

Crowther, R. A., Amos, L. A., Finch, J. T., DeRosier, D. J. & Klug, A. (1970a). Three dimensional reconstructions of spherical viruses by Fourier synthesis from electron micrographs. *Nature*, *226*, pp. 421-425.

Crowther, R. A., DeRosier, D. & Klug, A. (1970b). The Reconstruction of a Three-Dimensional Structure from Projections and its Application to Electron Microscopy. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *317*, pp. 319-340.

Crowther, R. A., Henderson, R. & Smith, J.M. (1996). MRC image processing programs. *J Struct Biol*, *116*, pp. 9-16.

Crowther, R. A., Kiselev, N. A., Böttcher, B., Berriman, J. A., Borisova, G. P., Ose, V. & Pumpens, P. (1994). Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell*, *77*, pp. 943-950.

DeRosier, D. J. & Klug, A. (1968). Reconstruction of three-dimensional structures from electron micrographs. *Nature*, *217*, pp. 130-134.

Dimmock, N. J., Easton, A. J. & Leppard, K. (2001). Introduction to Modern Virology. 5th Edition. Oxford, Blackwell Science.

Doan, D. N., Lee, K. C., Laurinmaki, P., Butcher, S., Wong, S. M. & Dokland, T. (2003). Three-dimensional reconstruction of hibiscus chlorotic ringspot virus. *J Struct Biol*, *144*, pp. 253-261.

Dubochet, J., Adrian, M., Lepault, J. & McDowall, A. (1985). Cryo-electron microscopy of vitrified biological specimens. *Trends Biochem. Sci*, *10*, pp. 143-146.

Erickson, H. P. & Klug, A. (1971). Measurement and Compensation of Defocusing and Aberrations by Fourier Processing of Electron Micrographs. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *261*, pp. 105-118.

Fabry, C. M. S., Rosa-Calatrava, M., Conway, J. F., Zubieta, C., Cusack, S., Ruigrok, R. W. H. & Schoehn, G. (2005). A quasi-atomic model of human adenovirus type 5 capsid. *EMBO J*, *24*, pp. 1645-1654.

Fischer, H., Polikarpov, I. & Craievich, A.F. (2004). Average protein density is a molecular-weight-dependent function. *Protein Science*, *13*, pp. 2825-2828.

Fokine, A., Kostyuchenko, V. A., Efimov, A. V., Kurochkina, L. P., Sykilinda, N. N., Robben, J., Volckaert, G., Hoenger, A., Chipman, P. R., Battisti, A. J., Rossmann, M. G. & Mesyanzhinov, V.V. (2005). A three-dimensional cryo-electron microscopy structure of the bacteriophage phiKZ head. *J Mol Biol*, *352*, pp. 117-124.

Frank, J. (2002). Single-particle imaging of macromoecules by cryo-electrom microscopy. *Annual Review of Biophysics and Biomolecular Structure*, *31*, pp. 303-319.

Frank, J. (2006). Three-dimensional Electron Microscopy of Macromolecular Assemblies Visualization of Biological Molecules in Their Native State. 2nd Edition. New York, Oxford University Press.

Frank, J. & Penczek, P. (1995). On the correction of the contrast transfer function in biological electron microscopy. *Optik(Stuttgart)*, *98*, pp. 125-129.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Ladjadj, M. & Leith, A. (1996). SPIDER and WEB: Processing and Visualization of Images in 3D Electron Microscopy and Related Fields. *J. Structural Biology*, *116*, pp. 190-199.

Fuller, S. D. (1987). The T= 4 envelope of Sindbis virus is organized by interactions with a complementary T= 3 capsid. *Cell*, *48*, pp. 923-934.

Fuller, S. D., Berriman, J. A., Butcher, S. J. & Gowen, B.E. (1995). Low pH induces swiveling of the glycoprotein heterodimers in the Semliki Forest virus spike complex. *Cell*, *81*, pp. 715-725.

Fuller, S. D., Butcher, S. J., Cheng, R. H. & Baker, T.S. (1996). Three-Dimensional Reconstruction of Icosahedral Particles - The Uncommon Line. *Journal of Structural Biology*, *116*, pp. 48-55.

Grimes, J. M., Jakana, J., Ghosh, M., Basak, A. K., Roy, P., Chiu, W., Stuart, D. I. & Prasad, B.V. (1997). An atomic model of the outer layer of the bluetongue virus core derived from X-ray crystallography and electron cryomicroscopy. *Structure*, *5*, pp. 885-893.

Hall, C. E. (1955). Electron Densitometry of Stained Virus Particles. *The Journal of Biophysical and Biochemical Cytology*, *1*, pp. 1-12.

Hartley, C. J., Greenwood, D. R., Gilbert, R. J. C., Masoumi, A., Gordon, K. H. J., Hanzlik, T. N., Fry, E. E., Stuart, D. I. & Scotti, P.D. (2005). Kelp fly virus: a novel group of insect picorna-like viruses as defined by genome sequence analysis and a distinctive virion structure. *J Virol*, *79*, pp. 13385-13398.

Henderson, R. (2004). Realizing the potential of electron cryo-microscopy. *Q Rev Biophys*, *37*, pp. 3-13.

Hewat, E. A. & Blaas, D. (1996). Structure of a neutralizing antibody bound bivalently to human rhinovirus 2. *EMBO J*, *15*, pp. 1515-1523.

Huiskonen, J. T., de Haas, F., Bubeck, D., Bamford, D. H., Fuller, S. D. & Butcher, S.J. (2006). Structure of the bacteriophage phi6 nucleocapsid suggests a mechanism for sequential RNA packaging. *Structure*, *14*, pp. 1039-1048.

Huiskonen, J. T., Kivelä, H. M., Bamford, D. H. & Butcher, S.J. (2004). The PM2 virion has a novel organization with an internal membrane and pentameric receptor binding spikes. *Nat Struct Mol Biol*, *11*, pp. 850-856.

Iwataki, M., Wong, M. & Fukuyo, Y. (2002). New record of Heterocapsa circularisquama(Dinophyceae) from Hong Kong. *Fisheries Science*, *68*, pp. 1161-1163.

Jäälinoja, H. T., Huiskonen, J. T. & Butcher, S.J. (2007). Electron cryomicroscopy comparison of the architectures of the enveloped bacteriophages phi6 and phi8. *Structure*, *15*, pp. 157-167.

Jiang, W. & Chiu, W. (2002). Web-based Simulation for Contrast Transfer Function and Envelope Functions. *Microscopy and Microanalysis*, *7*, pp. 329-334.

Jiang, W., Chang, J., Jakana, J., Weigele, P., King, J. & Chiu, W. (2006). Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature*, *439*, pp. 612-616.

Johnson, J. E. (1996). Functional implications of protein-protein interactions in icosahedral viruses. *Proc Natl Acad Sci U S A*, *93*, pp. 27-33.

Johnson, J. E. & Speir, J.A. (1997). Quasi-equivalent viruses: a paradigm for protein assemblies. *J. Mol. Biol*, *269*, p. 665–675.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, *292*, pp. 195-202.

Joyeux, L. & Penczek, P.A. (2002). Efficiency of 2D alignment methods. *Ultramicroscopy*, *92*, pp. 33-46.

Kitamura, N., Semler, B. L., Rothberg, P. G., Larsen, G. R., Adler, C. J., Dorner, A. J., Emini, E. A., Hanecak, R., Lee, J. J. & van der Werf, S. (1981). Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature*, *291*, pp. 547-553.

Klug, A. & Caspar, D.L. (1960). The structure of small viruses. *Adv Virus Res*, *7*, pp. 225-325.

Kovgan, A. A. & Zhdanov, V.M. (1989). [Structure and function of the genome of the cauliflower mosaic virus and its use in genetic engineering]. *Vopr Virusol*, *34*, pp. 523-533.

Krol, M. A., Olson, N. H., Tate, J., Johnson, J. E., Baker, T. S. & Ahlquist, P. (1999). RNA-controlled polymorphism in the in vivo assembly of 180-subunit and 120-subunit virions from a single capsid protein. *Proc Natl Acad Sci U S A*, *96*, pp. 13650-13655.

Kronenberg, S., Kleinschmidt, J. A. & Böttcher, B. (2001). Electron cryo-microscopy and image reconstruction of adeno-associated virus type 2 empty capsids. *EMBO Rep*, *2*, pp. 997-1002.

Lander, G. C., Tang, L., Casjens, S. R., Gilcrease, E. B., Prevelige, P., Poliakov, A., Potter, C. S., Carragher, B. & Johnson, J.E. (2006). The structure of an infectious P22 virion shows the signal for headful DNA packaging. *Science*, *312*, pp. 1791-1795.

Lebarron, J. (2005). *SPIDER: SY DOC (Create Symmetry Document file)*. Retrieved 5 October 2007 from http://www.wadsworth.org/spider_doc/spider/docs/man/sy.html.

Lee, K. K. & Johnson, J.E. (2003). Complementary approaches to structure determination of icosahedral viruses. *Curr Opin Struct Biol*, *13*, pp. 558-569.

Leiman, P. G., Battisti, A. J., Bowman, V. D., Stummeyer, K., Mühlenhoff, M., Gerardy-Schahn, R., Scholl, D. & Molineux, I.J. (2007). The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of new host specificities. *J Mol Biol, 371*, pp. 836-849.

Liu, X., Jiang, W., Jakana, J. & Chiu, W. (2007). Averaging tens to hundreds of icosahedral particle images to resolve protein secondary structure elements using a Multi-path Simulated Annealing optimization algorithm. *Journal of Structural Biology, 160*, pp. 11-27.

Ludtke, S. J., Baldwin, P. R. & Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol, 128*, pp. 82-97.

Ludtke, S. J., Jakana, J., Song, J. L., Chuang, D. T. & Chiu, W. (2001). A 11.5 A Single Particle Reconstruction of GroEL Using EMAN. *Journal of Molecular Biology, 314*, pp. 241-250.

Mancini, E. J., Clarke, M., Gowen, B. E., Rutten, T. & Fuller, S.D. (2000). Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus. *Mol Cell, 5*, pp. 255-266.

Mancini, E. J., de Haas, F. & Fuller, S.D. (1997). High-resolution icosahedral reconstruction: fulfilling the promise of cryo-electron microscopy. *Structure, 5*, pp. 741-750.

Matsuyama, Y. (1999). Harmful effect of dinoflagellate Heterocapsa circularisquama on shellfish aquaculture in Japan. *JARQ, 33*, pp. 283-293.

Metcalf, P., Cyrklaff, M. & Adrian, M. (1991). The three-dimensional structure of reovirus obtained by cryo-electron microscopy. *EMBO J, 10*, pp. 3129-3136.

Mindell, J. A. & Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol, 142*, pp. 334-347.

Miyazaki, Y., Nakashima, T., Iwashita, T., Fujita, T., Yamaguchi, K. & Oda, T. (2005). Purification and characterization of photosensitizing hemolytic toxin from harmful red tide phytoplankton, Heterocapsa circularisquama. *Aquatic Toxicology, 73*, pp. 382-393.

Mizumoto, H., Tomaru, Y., Takao, Y., Shirai, Y. & Nagasaki, K. (2007). Intraspecies host specificity of a single-stranded RNA virus infecting a marine photosynthetic protist is determined at the early steps of infection. *J Virol, 81*, pp. 1372-1378.

Morais, M. C., Choi, K. H., Koti, J. S., Chipman, P. R., Anderson, D. L. & Rossmann, M.G. (2005). Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29. *Mol Cell, 18*, pp. 149-159.

Mukhopadhyay, S., Chipman, P. R., Hong, E. M., Kuhn, R. J. & Rossmann, M.G. (2002). In vitro-assembled alphavirus core-like particles maintain a structure similar to that of nucleocapsid cores in mature virus. *J Virol, 76*, pp. 11128-11132.

Mukhopadhyay, S., Zhang, W., Gabler, S., Chipman, P. R., Strauss, E. G., Strauss, J. H., Baker, T. S., Kuhn, R. J. & Rossmann, M.G. (2006). Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses. *Structure, 14*, pp. 63-73.

Nagasaki K., Tomaru Y., Shirai Y., Takao Y., Mizumoto H. (2006). Dinoflagellate-infecting viruses. *Journal of the Marine Biological Association of the United Kingdom, 86*, pp. 469-474.

Nagasaki, K., Shirai, Y., Takao, Y., Mizumoto, H., Nishida, K. & Tomaru, Y. (2005). Comparison of genome sequences of single-stranded RNA viruses infecting the bivalve-killing dinoflagellate Heterocapsa circularisquama. *Appl Environ Microbiol*, *71*, pp. 8888-8894.

Nagasaki, K., Tomaru, Y., Nakanishi, K., Hata, N., Katanozaka, N. & Yamaguchi, M. (2004). Dynamics of Heterocapsa circularisquama (Dinophyceae) and its viruses in Ago Bay, Japan. *Aquatic Microbial Ecology*, *34*, pp. 219-226.

Ohi, M., Li, Y., Cheng, Y. & Walz, T. (2004). Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol Proced Online*, *6*, pp. 23-34.

Opalka, N., Tihova, M., Brugidou, C., Kumar, A., Beachy, R. N., Fauquet, C. M. & Yeager, M. (2000). Structure of native and expanded sobemoviruses by electron cryo-microscopy and image reconstruction. *J Mol Biol*, *303*, pp. 197-211.

Penczek, P. A., Renka, R. & Schomberg, H. (2004). Gridding-based direct Fourier inversion of the three-dimensional ray transform. *Journal of the Optical Society of America A*, *21*, pp. 499-509.

Penczek, P. A., Zhu, J. & Frank, J. (1996). A common-lines based method for determining orientations for N> 3 particle projections simultaneously. *Ultramicroscopy*, *63*, pp. 205-218.

Penczek, P. A., Zhu, J., Schröder, R. & Frank, J. (1997). Three dimensional reconstruction with contrast transfer compensation from defocus series. *Scanning Microscopy*, *11*, pp. 147-154.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T.E. (2004). UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.*, *25*, pp. 1605-1612.

Prasad, B. V. & Prevelige Jr, P.E. (2003). Viral genome organization. *Adv Protein Chem*, *64*, pp. 219-258.

Prasad, B. V. V., Burns, J. W., Marietta, E., Estes, M. K. & Chiu, W. (1990). Localization of VP 4 neutralization sites in rotavirus by three-dimensional cryo-electron microscopy. *Nature*, *343*, pp. 476-479.

Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. (1987). Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *J Microsc*, *146*, pp. 113-136.

Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C. & Harrison, S.C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. *Nature*, *375*, pp. 291-298.

Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E., Johnson, J. E., Douglas, T. & Young, M. (2004). The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc Natl Acad Sci U S A*, *101*, pp. 7716-7720.

Roseman, A. M., Berriman, J. A., Wynne, S. A., Butler, P. J. G. & Crowther, R.A. (2005). A structural model for maturation of the hepatitis B virus core. *Proc Natl Acad Sci U S A*, *102*, pp. 15821-15826.

Rossmann, M. G. & Johnson, E. (1989). Icosahedral RNA virus structure. *Annu. Rev. Biochem*, *58*, pp. 533-573.

Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H. J., Johnson, J. E., Kamer, G., Luo, M. & Mosser, A.G. (1985). Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature*, *317*, pp. 145-153.

Saad, A., Ludtke, S. J., Jakana, J., Rixon, F. J., Tsuruta, H. & Chiu, W. (2001). Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination. *J. Struct. Biol*, *133*, pp. 32-42.

Saban, S. D., Silvestry, M., Nemerow, G. R. & Stewart, P.L. (2006). Visualization of alpha-helices in a 6-angstrom resolution cryoelectron microscopy structure of adenovirus allows refinement of capsid protein assignments. *J Virol*, *80*, pp. 12049-12059.

Saibil, H. R. (2000). Macromolecular structure determination by cryo-electron microscopy. *Acta Crystallogr D Biol Crystallogr*, *56*, pp. 1215-1222.

Saugar, I., Luque, D., Oña, A., Rodríguez, J. F., Carrascosa, J. L., Trus, B. L. & Castón, J.R. (2005). Structural polymorphism of the major capsid protein of a double-stranded RNA virus: an amphipathic alpha helix as a molecular switch. *Structure*, *13*, pp. 1007-1017.

Schoehn, G., Moss, S. R., Nuttall, P. A. & Hewat, E.A. (1997). Structure of Broadhaven virus by cryoelectron microscopy: correlation of structural and antigenic properties of Broadhaven virus and bluetongue virus outer capsid proteins. *Virology*, *235*, pp. 191-200.

Sellner, K. G., Doucette G J & Kirkpatrick G J (2000). Harmful algal blooms: causes, impacts and detection. *Journal of Industrial Microbiology and Biotechnology*, *30*, pp. 383-406.

Shannon, C. E. (1949). Communication in the presence of noise. *Proc. IRE*, *37*, p. 10–21.

Shepherd, C. M., Borelli, I. A., Lander, G., Natarajan, P., Siddavanahalli, V., Bajaj, C., Johnson, J. E., Brooks, C. L. 3. & Reddy V S. (2006). VIPERdb: a relational database for structural virology. *Nucleic Acids Res.*, *34 (Database Issue)*, p. D386-D389.

Smith, M. F. & Langmore, J.P. (1992). Quantitation of molecular densities by cryo-electron microscopy. Determination of the radial density distribution of tobacco mosaic virus. *J Mol Biol*, *226*, pp. 763-774.

Speir, J. A., Munshi, S., Wang, G., Baker, T. S. & Johnson, J.E. (1995). Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure*, *3*, pp. 63-78.

Stewart, P. L., Burnett, R. M., Cyrklaff, M. & Fuller, S.D. (1991). Image reconstruction reveals the complex molecular organization of adenovirus. *Cell*, *67*, pp. 145-154.

Tagari, M., Newman, R., Chagoyen, M., Carazo, J. M. & Henrick, K. (2002). New electron microscopy database and deposition system. *Trends Biochem Sci*, *27*, p. 589.

Tang, L., Johnson, K. N., Ball, L. A., Lin, T., Yeager, M. & Johnson, J.E. (2001). The structure of pariacoto virus reveals a dodecahedral cage of duplex RNA. *Nat Struct Biol*, *8*, pp. 77-83.

Tang, L., Lin, C., Krishna, N. K., Yeager, M., Schneemann, A. & Johnson, J.E. (2002). Virus-like particles of a fish nodavirus display a capsid subunit domain organization different from that of insect nodaviruses. *J Virol*, *76*, pp. 6370-6375.

Tarutani, K., Nagasaki, K., Itakura, S. & Yamaguchi, M. (2001). Isolation of a virus infecting the novel shellfish-killing dinoflagellate Heterocapsa circularisquama . *Aquatic Microbial Ecology*, *23*, pp. 103-111.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, *25*, pp. 4876-4882.

Thon, F. (1966). Zur Defokussierungsabhängigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung. *Z. Naturforsch*, *219*, pp. 179-180.

Thuman-Commike, P. A. (2001). Single particle macromolecular structure determination via electron microscopy. *FEBS Letters*, *505*, pp. 199-205.

Thuman-Commike, P. A. & Chiu, W. (2000). Reconstruction principles of icosahedral virus structure determination using electron cryomicroscopy. *Micron*, *31*, pp. 687-711.

Tomaru, Y., Hata, N., Masuda, T., Tsuji, M., Igata, K., Masuda, Y., Yamatogi, T., Sakaguchi, M. & Nagasaki, K. (2007). Ecological dynamics of the bivalve-killing dinoflagellate Heterocapsa circularisquama and its infectious viruses in different locations of western Japan. *Environ Microbiol*, *9*, pp. 1376-1383.

Tomaru, Y., Katanozaka, N., Nishida, K., Shirai, Y., Tarutani, K., Yamaguchi, M. & Nagasaki, K. (2004). Isolation and characterization of two distinct types of HcRNAV, a single-stranded RNA virus infecting the bivalve-killing microalgae Heterocapsa circularisquama . *Aquatic Microbial Ecology*, *34*, pp. 207-218.

Trus, B. L., Roden, R. B., Greenstone, H. L., Vrhel, M., Schiller, J. T. & Booy, F.P. (1997). Novel structural features of bovine papillomavirus capsid revealed by a three-dimensional reconstruction to 9 A resolution. *Nat Struct Biol*, *4*, pp. 413-420.

van Heel, M. & Frank, J. (1981). Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy*, *6*, pp. 187-194.

van Heel, M., Gowen, B., Matadeen, R., Orlova, E. V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R. & Schatz, M. (2000). Single-particle electron cryo-microscopy: towards atomic resolution. *Quarterly Reviews of Biophysics*, *33*, pp. 307-369.

Vogel, R. H., Provencher, S. W., von Bonsdorff, C. H., Adrian, M. & Dubochet, J. (1986). Envelope structure of Semliki Forest virus reconstructed from cryo-electron micrographs. *Nature*, *320*, pp. 533-535.

Wery, J. P., Reddy, V. S., Hosur, M. V. & Johnson, J.E. (1994). The refined three-dimensional structure of an insect virus at 2.8 A resolution. *J Mol Biol*, *235*, pp. 565-586.

Xiang, Y., Morais, M. C., Battisti, A. J., Grimes, S., Jardine, P. J., Anderson, D. L. & Rossmann, M.G. (2006). Structural changes of bacteriophage phi29 upon DNA packaging and release. *EMBO J*, *25*, pp. 5229-5239.

Zhang, W., Fisher, B. R., Olson, N. H., Strauss, J. H., Kuhn, R. J. & Baker, T.S. (2002). Aura virus structure suggests that the T=4 organization is a fundamental property of viral structural proteins. *J Virol*, *76*, pp. 7239-7246.

Zhang, Y., Corver, J., Chipman, P. R., Zhang, W., Pletnev, S. V., Sedlak, D., Baker, T. S., Strauss, J. H., Kuhn, R. J. & Rossmann, M.G. (2003). Structures of immature flavivirus particles. *EMBO J*, *22*, pp. 2604-2613.

Zhang, Y., Kaufmann, B., Chipman, P. R., Kuhn, R. J. & Rossmann, M.G. (2007). Structure of immature West Nile virus. *J Virol*, *81*, pp. 6141-6145.

Zhou, Z. H. & Chiu, W. (2003). Determination of icosahedral virus structures by electron cryomicroscopy at subnanometer resolution. In F. M. Richards, D. S. Eisenberg & J. Kuriyan (Eds.), *Advances in protein chemistry*. Amsterdam: Academic Press. pp. 93-124.

Zhou, Z. H., Baker, M. L., Jiang, W., Dougherty, M., Jakana, J., Dong, G., Lu, G. & Chiu, W. (2001). Electron cryomicroscopy and bioinformatics suggest protein fold models for rice dwarf virus. *Nat Struct Biol*, *8*, pp. 868-873.

Zhou, Z. H., Dougherty, M., Jakana, J., He, J., Rixon, F. J. & Chiu, W. (2000). Seeing the herpesvirus capsid at 8.5 A. *Science*, *288*, pp. 877-880.

Zhou, Z. H., Prasad, B. V., Jakana, J., Rixon, F. J. & Chiu, W. (1994). Protein subunit structures in the herpes simplex virus A-capsid determined from 400 kV spot-scan electron cryomicroscopy. *J Mol Biol*, *242*, pp. 456-469.

Zhu, J., Penczek, P. A., Schröder, R. & Frank, J. (1997). Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S Escherichia coli ribosome. *J. Struct. Biol*, *118*, pp. 197-219.

Zlotnick, A. & Stray, S.J. (2003). How does your virus grow? Understanding and interfering with virus assembly. *Trends Biotechnol*, *21*, pp. 536-542.

Appendix A

Table A: *List of defocii of the micrographs*

Micrograph	Defocus (µm)	Micrograph	Defocus (µm)
302	31146.36	495	37967.63
303	44046.94	496	35422.41
304	42296.29	497	36278.10
305	32720.93	498	39925.94
306	34588.38	499	34837.26
307	31642.18	501	37718.24
308	23626.57	502	37310.70
309	35364.70	504	31393.72
310	23886.59	505	32212.15
311	38111.93	506	33371.91
313	28324.11	519	37425.11
314	32468.67	521	41066.31
315	35666.89	522	38041.47
316	35978.95	524	36866.78
317	33486.74	527	35049.51
318	33271.67	529	37751.71
319	29682.29	530	36671.78
321	34739.64	531	32701.49
322	42997.44	532	30999.58
323	28399.59	534	33111.73
324	37196.00	536	31245.15
325	57817.96	539	29865.60
467	41205.11	546	29896.73
471	31197.15	547	34767.30
473	30075.05	551	30437.68
477	28126.47	553	32522.78
479	30541.77	554	31869.65
480	33057.11	555	16370.61
481	32227.96	557	33630.89
482	30507.16	558	48891.95
483	30441.89	559	31756.26
484	30917.20	560	44332.38
486	31788.57	561	32045.41
487	30573.84	562	43134.52
488	31451.80	563	32010.50
489	29596.16	565	34018.06
490	27396.60		

Appendix B

Table B: *Icosahedral symmetry file for MRC orientation of an icosahedral particle.*

	ψ	θ	φ			ψ	θ	φ	
1	3	0.0000	0.0000	0.0000	31	3	-58.283	36.000	121.72
2	3	-58.283	72.000	-121.72	32	3	-159.09	60.000	-159.09
3	3	-58.283	144.00	-121.72	33	3	159.09	120.00	-159.09
4	3	121.72	144.00	58.283	34	3	58.283	144.00	121.72
5	3	121.72	72.000	58.283	35	3	0.0000	90.000	90.000
6	3	-121.72	36.000	-121.72	36	3	-58.283	108.00	121.72
7	3	-121.72	108.00	-121.72	37	3	-121.72	72.000	121.72
8	3	0.0000	180.00	0.0000	38	3	159.09	60.000	159.09
9	3	58.283	108.00	58.283	39	3	90.000	90.000	-180.00
10	3	58.283	36.000	58.283	40	3	20.905	120.00	159.09
11	3	-90.000	90.000	-180.00	41	3	0.0000	180.00	0.0000
12	3	-159.09	120.00	159.09	42	3	-121.72	108.00	58.283
13	3	121.72	108.00	121.72	43	3	-121.72	36.000	58.283
14	3	58.283	72.000	121.72	44	3	58.283	36.000	-121.72
15	3	-20.905	60.000	159.09	45	3	58.283	108.00	-121.72
16	3	-121.72	144.00	121.72	46	3	121.72	108.00	-58.283
17	3	180.00	90.000	90.000	47	3	-159.09	120.00	-20.905
18	3	121.72	36.000	121.72	48	3	-90.000	90.000	0.0000
19	3	20.905	60.000	-159.09	49	3	-20.905	60.000	-20.905
20	3	-20.905	120.00	-159.09	50	3	58.283	72.000	-58.283
21	3	159.09	120.00	20.905	51	3	121.72	36.000	-58.283
22	3	-159.09	60.000	20.905	52	3	180.00	90.000	-90.000
23	3	-58.283	36.000	-58.283	53	3	-121.72	144.00	-58.283
24	3	0.0000	90.000	-90.000	54	3	-20.905	120.00	20.905
25	3	58.283	144.00	-58.283	55	3	20.905	60.000	20.905
26	3	159.09	60.000	-20.905	56	3	0.0000	0.0000	0.0000
27	3	-121.72	72.000	-58.283	57	3	121.72	72.000	-121.72
28	3	-58.283	108.00	-58.283	58	3	121.72	144.00	-121.72
29	3	20.905	120.00	-20.905	59	3	-58.283	144.00	58.283
30	3	90.000	90.000	0.0000	60	3	-58.283	72.000	58.283